Hierarchical Control in the Execution of Action Sequences: Tests of Two Invariance Properties

Saul Sternberg, Ronald L. Knoll, & David L. Turock

1 Hierarchical Control in the Execution of Action Sequences: Tests of Two Invariance Properties

Saul Sternberg
University of Pennsylvania

Ronald L. Knoll
AT&T Bell Laboratories

David L. Turock
Bell Communications Research

ABSTRACT

What might it mean for execution of an action sequence to be controlled hierarchically? We argue that if production of a sequence consists of the execution of nested constituent subsequences, then it should be characterized by two invariance properties—properties that limit the effects of one part of the sequence on another. Because one such constituent structure merely partitions the stream of action into *action units*, these properties have wide applicability. According to *low-level invariance*, the process that executes a constituent should not be influenced by changes in any higher level constituent. According to *high-level invariance*, changes in a constituent should have at most limited and local effects on higher level constituents. We report on tests of these two properties in the rapid production of brief utterances and short strings of keystrokes, in which we examine the effects of sequence length, serial position, and unit size on measures of timing. The tests support the existence of hierarchical constituents at the level of the stroke in typing and the stress group in speech, but provide only limited evidence for deeper hierarchical structure.

INTRODUCTION

In this chapter we investigate one sense in which execution of action sequences might be hierarchical, by virtue of their being composed of separately controlled constituent subsequences. We argue that to claim merely that a continuous stream of action can be decomposed into a string of

3

concatenated *units* is to assert a simple form of such hierarchical structure; properties that distinguish the constituents in such a structure should thus characterize purported action units. We focus on two invariance properties that flow from this sense of hierarchical control—properties that reflect the idea that different control levels function autonomously and thus impose limits on the effects of sequential context—and report attempts to determine the extent to which these properties characterize timing in the rapid production of speech and keystroke sequences. The spirit of our inquiry is to minimize the number of ancillary assumptions and thus avoid strong models. Our aim is to illustrate some alternative approaches to testing for hierarchical structure, applying them in most cases to data collected for other purposes.

In our examination of evidence we consider instances where short action sequences prescribed well in advance are correctly produced under time pressure, and where time measurements thus indicate performance constraints. We justify this choice of procedure by our desire to separate the execution of planned sequences from the process by which they are planned. Performance measurements indicate that under these conditions a plan or "program" for the entire sequence exists before it is initiated (Rosenbaum, Kenny, & Derr, 1983; Sternberg, Monsell, Knoll, & Wright, 1978). We deliberately do not investigate cases where choice of sequence is free rather than prescribed (e.g., Fentress, 1983), or where errors are of central interest (e.g., Shattuck-Hufnagel, 1983): The choice of action element (when choice is free) or errors in such choice (when the sequence is prescribed) could occur during the planning process as well as the execution process, whereas temporal effects in the execution of prescribed sequences seem more likely to be associated with the latter. Because our concern is with how the control mechanism selects successive actions, we consider primarily sequences whose successive actions are distinct rather than repeating.

PROPERTIES OF SEQUENCES UNDER HIERARCHICAL CONTROL

Concepts of Hierarchy

One example of the numerous ways[1] in which the term *hierarchy* has been used is to denote a simple ordering on some dimension, often described as a set of *levels*. More interesting are tree-like branching structures, con-

---

[1]See, for example, Dawkins (1976); Estes (1972); Gallistel (1980); Johnson (1970); Lehrdahl and Jackendoff (1983); Narmour (1983); Nelson (1973); and Wilson (1969).

sisting of a set of elements (*nodes*) at different levels, partially ordered by a relation (*branches*), usually antisymmetric and transitive (Wall, 1972). Examples of relations are *inclusion* and *control*. If the relation is inclusion, we have a classification hierarchy in which each class (at one level) consists of a set of subclasses (at the next), and a subclass can belong to only one class. If the relation is control, each element (at one level) controls a set of elements (at the next), and an element can be controlled by at most one higher level element.

In applying this idea to action sequences we are interested in strings of rapid actions that can be said to contain one or more separately controlled substring *constituents* (or *units*), larger than a "single" action (that is, susceptible to further analysis), but smaller than the whole string (Gallistel, 1980, pp. 288–290). The validity of such an assertion hinges theoretically on the specification of criteria that define a substring as a constituent, and empirically on the demonstration that such criteria are satisfied. This sense of hierarchy should be distinguished from the idea of different levels of specificity, such that more detailed aspects of the same action are controlled autonomously at lower levels, closer to its execution (Szentagothai & Arbib, 1975), and where control branches can converge and cross (Gallistel, 1980).

Perhaps the most familiar domain where hierarchically organized sequences have been used to characterize human behavior is language. The constituent structure of sentences is represented in linguistics by the recursively branching phrase marker (Wall, 1972), which we introduce to add precision to the idea of hierarchical structure. The phrase marker can be expressed as either an *ordered tree*, (Fig. 1.1A), or a *labeled-bracket structure* (Fig. 1.1B). An ordered tree consists of a set of *nodes* (from a *root* at the highest level to *leaves* at the lowest, or first) connected by diverging and non-crossing *branches*. Any pair of nodes is related either by *dominance* (for instance, in Fig. 1.1 *VP* dominates the right-hand *N*, and is thus at a higher "level") or by *precedence* (for instance, the upper *NP* precedes *V*) but not by both. Each node corresponds to a substring *constituent* of the sentence. The nesting of constituents is more evident in the labeled-bracket structure, in which dominance becomes inclusion.

A hierarchically organized string may thus be defined as a set of constituent (sub)strings, partially ordered by inclusion, such that any string either fully contains, or is fully contained in, or is disjoint from any other string, and such that disjoint constituents are temporally ordered. A string of "action unit" substrings defines the most shallow structure that can be described as hierarchical; in a deeper hierarchy the substrings would be further partitioned into smaller disjoint substrings.

FIG. 1.1.   Two forms of the phrase marker representation of a sentence. A: An ordered tree composed of nodes and branches. B: A partially equivalent structure of nested labeled brackets.

Two Invariance Properties of Hierarchically Controlled Sequences

Augmenting of hierarchical structure with hierarchical control.   The phrase marker is a hierarchical structure with no commitment to any particular embodiment in real time or to any specification of the flow of control. To convert such a structure into one that can actually control the execution of the sequence it describes, we must augment it with an interpretive

process (Greeno & Simon, 1974) that uses or operates on the structure to create an action sequence. In a plausible process of this kind, the nodes, which correspond to nested constituents, also correspond to loci of control. Let us use *program* to refer to the structure that produces the entire sequence when the process is applied to it. Then a constituent subsequence is produced by invoking a *subprogram* that produces the subsequence and is associated with the constituent's node; such actuating of subprograms proceeds recursively until first-level nodes are reached.

In the present chapter we focus on the implications of independence between control levels and the resulting constraints on context effects. Such implications are especially clear if the process that operates on the structure does so by *transfer of control* from program (at one level) to subprogram (at a lower level) and back.[2] Production of a constituent results from execution of a (sub)program that, we suggest, should be insensitive to properties of the superordinate program that invoked it. That is, once initiated, the subprogram is executed autonomously until completion, without reference to information stored at a higher level; control is returned to that level upon completion. Similarly, the superordinate program is insensitive to details of the subsequence; it "knows" at most whether or not the subprogram has completed its job.

Low-level and high-level invariance as properties of hierarchical control. Two consequences flow from these considerations. The first is a property we term *low-level invariance*: The process that executes a constituent substring should be invariant relative to changes in disjoint or higher level parts of the control structure. (When these two loci of change are distinguishable we have two kinds of low-level invariance: For example, if we altered only the content of a disjoint constituent we could test for invariance relative to a disjoint part of the control structure; if we altered the number of constituents we could test for invariance relative to a higher level part of the control structure.) Given low-level invariance, performance characteristics of all elements within a constituent should thus depend only on the content of that constituent; they should not be influenced by changes in the constituents of which it is a substring. An analogy in language arises in context-free phrase-structure grammars, in which the rewriting rule according to which a node is elaborated downward depends only on

---

[2]Tree-traversal models of sequence production (Collard & Povel, 1982; Greeno & Simon, 1974; Rosenbaum, Kenny, & Derr, 1983; Volpert, 1982) exemplify processes with transfer of control, processes that are natural in single-processor systems and supported by some of our analyses. An alternative (command/signal transmission) that is possible in multiprocessor systems and favored by one of our analyses is considered later.

the label of that node, and not on any other nodes. Low-level invariance, or an approximation thereof, is inherent in the idea of a *unit,* an entity that must preserve some or all of its properties across different contexts.

The second consequence is *high-level invariance:* relative to changes in a constituent substring, the higher level process that controls its execution along with the execution of disjoint other constituents, and hence those other constituents themselves, should be invariant. Changes in the content of a constituent should thus have only local effects (limited to the constituent itself) on the higher level constituents of which it is a substring. An analogy in language is the existence of syntactic dependence, such as number agreement between subject and verb, that is unaltered by expanding an intervening constituent and thus increasing their separation.

It might be argued that these invariance criteria are too stringent; the hypothesis of hierarchical structure should perhaps not be rejected in every instance in which one or the other criterion is found wanting. We might instead ask to what degree of approximation these properties obtain, and for which aspects of performance, in cases where we have other reasons to argue for hierarchical structure. Thus, Simon (1962) suggests that hierarchical systems might be "nearly decomposable," rather than displaying full invariance. On the other hand, it seems reasonable to start with the expectation of full invariance, and discover how far it will carry us.

The testing of invariance.   To test either form of invariance one assesses the effects of manipulating an experimental *factor* on some *measure.* To provide an interesting test, the factor should influence *some* aspect of the performance under study. For this reason we refer to a factor used in this way as an *influencer.* Although the invariance properties apply to any aspect of performance, our tests to date are limited to duration measures. To emphasize this limitation we sometimes use the term *duration invariance.*

Command/signal transmission as an alternative to transfer of control.   Strict interpretation of "transfer of control" requires that there be no temporal overlap of operations associated with different levels (Broadbent, 1977). Durations of operations at different levels must then contribute additively to sequence duration, and time measurements become especially revealing. Given such a process, however, we must reconcile the apparent continuity of action not only with the discreteness of the action unit, discussed later, but also with the possibility of time-consuming operations at high levels, which (on this account) should lead to epochs during which action is frozen. It may thus be necessary to at least augment a process of transfer of control such that action controlled by a low-level subprogram can outlast the execution of that subprogram, filling the gap during which

the next subprogram is being selected and invoked, but without delaying its execution (Sternberg, Knoll, Monsell, & Wright, 1989, Section 4.3).

These considerations, together with aspects of coordination of the hands in typewriting described later, also suggest that we consider a fundamentally different class of control processes that might operate on a hierarchical structure, in which transfer of control is replaced by *command/signal transmission:* A program invokes a subprogram by transmitting a command downward rather than relinquishing control; completion (or forthcoming completion) may be indicated by an upward signal, but return of control is not needed. Processes at different levels, or even at the same level, may thus overlap in time and proceed concurrently. Our invariance properties should still obtain, but their consequences may be obscured, depending on the extent of temporal overlap and other details; the interpretation of time measurements may be more complicated. If higher level operations must await a signal of completion from below, such a process could mimic transfer of control.

Fundamental Discreteness of Continuous Action

Action is in general continuous; even when it is punctuated by a series of distinctive time markers, such as keystroke actuation points in typing or articulatory closures in speech, the system is unlikely to be static during intervals associated with these events, nor do we wish to limit our analyses to those cases where it might be. Given such continuity, it is well to bring out a common but often implicit assumption: The apparently continuous action stream is fundamentally discrete. Without this assumption we could not readily consider the planning of action sequences, and we would not be attempting to identify action units. The idea is commonplace in thinking about speech production, which is often thought to have phonemes as low-level units, with their concatenation smoothed by coarticulation (Fowler, 1985; Perkell, 1980). In relation to manual movement trajectories, especially in handwriting, Morasso and others have shown how an inherently discrete and discontinuous mechanism can produce smooth movements by a particular form of coarticulation: the partial time overlap of consecutive units, achieved by superposition of their kinematic characterizations.[3] If such smoothing were produced by a process that operated on the output of a hierarchical control mechanism, then the resulting perturbations at constituent boundaries might obscure instances of expected invariance; If

---

[3]See Morasso (1983) and Morasso, Mussa Ivaldi, and Ruggiero (1983). In an experimental test using planar arm-movements, such superposition failed to describe the resultant trajectory (Massey, Schwartz, & Georgopoulos, 1986).

such effects were widespread it would be difficult to proceed further. We encounter cases in both speech and typing where we observe such contamination, but fortunately find it to be limited.

Alternative Criteria for Action Units

The decomposition of a continuous stream of action into units is an assertion of a particular hierarchical structure. Any criteria that might be diagnostic for action units are thus likely to be helpful in relation to constituents at any level in a hierarchical structure, and vice versa; this is especially evident when one recognizes that units at one level of description of a sequence are also likely to be both compounds at a lower level and subunits at a higher level.

Unit criteria that have been discussed or implied in previous work should be considered in relation to those here under investigation.[4] Ultimately we may discover that several of these properties tend to co-occur. Some are as follows:

1.   Stereotypy of unit execution. Execution of a unit should be invariant in the presence of contextual variation (as in low-level invariance) or from occasion to occasion. For example, time intervals between its parts or ratios of such intervals should be relatively constant. (See Barlow, 1977; Gentner, 1987; Terzuolo & Viviani, 1980.) A related criterion arises from Vorberg and Hambuch's (1978, 1984) definition of hierarchical timing discussed later: Time intervals between features within a unit should display smaller variances than intervals (of approximately the same mean duration) between features in different units.

2.   Correlation of component durations. Durations of components of the same unit, but not of different units,should be correlated over repeated productions (Benecke, Rothwell, Day, Dick, & Marsden, 1986).

3.   Temporal separation. Intervals separating successive component actions should be longer across than within unit boundaries (Gee & Grosjean, 1983; Machlis, 1977; Reitman & Reuter, 1980; Rosenbaum, et al., 1983).

---

[4]Given an explicit theory of the production of action sequences, definition of a unit or constituent may emerge naturally. For example, in the Rumelhart and Norman (1982) model of transcription typing, the parser transforms a long string of input characters into a sequence of substring constituents. And in the subprogram selection model developed for the production of short action sequences in speech and typing by Sternberg, Monsell, et al. (1978; see also Sternberg, Wright, Knoll, & Monsell, 1980; Sternberg, Knoll, Monsell, & Wright, 1983; and Sternberg, Knoll, et al., 1989) a unit at the level of the action sequence is defined as that segment of action controlled by a single subprogram, or associated with a single rendition of the subprogram-selection process.

4.   Equivalence of effects of units on other actions. Any effect on the remainder of a sequence of adding a unit should be independent of the size or content of that unit (Sternberg, Knoll, et al., 1989). Note that this is equivalent to high-level invariance, according to which one constituent can be replaced by another, with no effect on the remainder of the control structure.[5]

5.   Equivalence of effects of other actions on units. An effect *on* a unit of lengthening a sequence should be independent of the content of that unit (Sternberg, Knoll, et al., 1989). Note that although this requirement appears at first glance to be equivalent to low-level invariance, it is weaker: Low-level invariance requires *no* effect of changing the embedding structure, rather than merely an invariant effect.

6.   Preferential influence. Segments within the same unit should influence each other more than segments of different units. Fowler (1981, 1985) uses this criterion to support the idea of a stress-group unit in speech production. In a perceptual analog, Palmer (1977) proposes that elements within units of a hierarchically organized percept have relatively more influence on each other. Simon (1962) describes hierarchical systems as "nearly decomposable," displaying stronger interactions within than between units.

7.   Internal coherence. Parts of a unit should be difficult to separate. Segments within the same unit should thus never (or seldom) participate independently in errors of omission or misordering (Drummond, 1981; Fentress, 1983; Fromkin, 1981; Gallistel, 1980), and it should be more difficult to produce a prescribed rearrangement of a prepared sequence that breaks up units than one that leaves them intact (Gordon & Meyer, 1987).

8.   Process discreteness. A unit is a segment of performance that results from a single rendition of some process. This criterion is, of course, more model-dependent than the others because in general, a process is hypothetical and requires a model to be related to observables (Sternberg, Knoll, et al., 1989, Section 2.5).

9.   Bell-shaped velocity profile. In arm and hand movements, a unit can be distinguished by a bell-shaped velocity profile, and is thus a segment of the trajectory delimited by velocity minima (Abend, Bizzi, & Morasso, 1982).

---

[5]The equivalence can be seen as follows: Suppose equal effects on some univariate property P_a of sequence a, of adding (concatenating or incorporating) constituents u_x and u_y. That is, $P_a(a+u_x) - P_a(a) = P_a(a+u_y) - P_a(a)$. It follows that $P_a(a+u_x) = P_a(a+u_y)$, so that replacing u_x by u_y leaves P_a invariant. Conversely, it can be seen that invariance under replacement implies equal effects.

Existing Evidence for Hierarchical Control
in Sequence Production

The idea that fluently performed action sequences tend to be structured hierarchically, and that skill learning includes the creation of higher-level action units, dates back at least to the turn of the century (Book, 1908). However, there is a disparity between the enthusiasm for the hypothesis of hierarchical organization in the skilled execution of action sequences (e.g., Gallistel, 1980, Ch. 12; Johnson, 1970; Keele, 1987; MacKay, 1982; Miller, Galanter, & Pribram, 1960) and the persuasiveness of much of the cited evidence. For example, contrary to frequent citation, Bryan and Harter's (1899) claims were for hierarchical structure in receiving, rather than sending Morse code. Other evidence does not exclude a perceptual locus (Leonard & Newman, 1964). The often-cited conclusions of Book (1908) about multiple-stroke units in typewriting depend exclusively on introspective evidence. Furthermore, they have been only partially confirmed by the effects on transcription speed at different skill levels of the length of familiar sequences incorporated in the copy (Fendrick, 1937; West, 1969, Ch. 3).[6] More recently, some evidence consistent with hierarchical structure of the representation that controls the execution of action sequences has begun to emerge: Gordon and Meyer (1987) have applied the internal coherence criterion listed in the previous section to speech production, and Rosenbaum et al. (1983), whose work we discuss later, have applied the temporal separation criterion to finger tapping (but see also Klein, 1983, and Rosenbaum, 1983).

Because their problem is the timing of "identical" actions rather than the choice of distinct ones, Vorberg and Hambuch's (1978, 1984) attempts to determine conditions (if any) for hierarchical control in the timing of repeated rhythmic patterns fall outside the domain of central interest here; they are nonetheless important in any review. Suppose a series of taps is produced under instructions to make them approximately isochronic but to group them in triple meter. Denote the times of occurrence of each of the successive taps as

$$t_{0,0}, t_{0,1}, t_{0,2}, t_{1,0}, t_{1,1}, t_{1,2}, \ldots, t_{n,0}, t_{n,1}, t_{n,2}, \ldots$$

According to a serial-timing model, each tap is timed from the preceding one. In a typical hierarchical-timing model, the timing of one or more

---

[6]Inferences in these studies about the size of the action unit depend on two questionable assumptions: First, skill level is assumed to have no influence on the perceptual or translation processes in transcription; second, given no change in unit size, skill level is assumed to have a multiplicative effect on transcription speed for different kinds of copy.

intervals [e.g., $(t_{n,0}, t_{n,1})$ and $(t_{n,1}, t_{n,2})$] can be nested within the timing of a longer interval between nonsuccessive events [e.g., $(t_{n,0}, t_{n+1,0})$] such that long-interval and short-interval "timers" can start in synchrony and operate in parallel. If so, the durations of one or more (short) intervals [$(t_{n,2}, t_{n+1,0})$ in our example] are determined indirectly, as the difference between other, directly timed intervals.

As shown by Vorberg and Hambuch's elegant analyses, numerous properties of the resultant time intervals distinguish these two models; perhaps the most obvious is *phase-dependent variability*: For serial timing, the variance, $s_i^2 = Var_n(T_{n,i})$ of the intervals $\{T_{n,i}\} = \{t_{n+1,i} - t_{n,i}\}$ between corresponding taps is independent of i, because each such interval is generated by a sum of intervals from the same three timers. In contrast, for hierarchical timing this variance depends systematically on i, because for $i = 0$ it is generated by one timer whereas for $i = 1$ or $i = 2$ it is generated by a linear combination of intervals from three timers. Under the instructions described above, subjects show clear evidence of grouping in the periodicity of the mean intervals $\{t_{.,i+1} - t_{.,i}\}$ between successive taps. This might tempt one to conclude in favor of a hierarchical structure, but the analysis of patterns of variance and covariance instead indicates serial timing. Under instructions to produce taps in an uneven rhythm, subjects' phase-dependent variance provides evidence against serial timing, but their covariance data argue against hierarchical timing. Like evidence for hierarchical structure in the selection of actions, the evidence for hierarchical timing has thus been elusive.[7]

Patterns of error have sometimes been cited as evidence for hierarchical structure and for the existence of particular units, especially in speech production (see Dell, 1984; Fromkin, 1971, 1981; Shattuck-Hufnagel, 1983). In evaluating such evidence, however, the distinction should be made between the idea that *some* representation during the evolution of a movement pattern contains intermediate-size subsequence units, and the idea that this is true of the one finally used in executing the pattern. Given separability of planning and execution processes, it is important to distinguish between *units of planning* and *units of execution* (Monsell & Sternberg, 1981; Rosenbaum, et al., 1983). Errors may arise during planning or execution. Our hope is that *temporal* aspects of the execution of a sequence prescribed in advance are more likely than errors to reflect the representation used during its execution. It is partly for this reason that despite the generality of the two invariance properties under consideration,

---

[7]Jagacinski, Marshburn, Klapp, and Jones (1988) have recently produced an analysis that favors a modified version of Vorberg and Hambuch's hierarchical model, in which concurrently operating timers produce positively correlated (rather than independent) intervals.

our tests thus far are restricted to the *duration* invariance of elements and sequences.

TESTS OF LOW-LEVEL INVARIANCE

Existence of an Action Unit in Rapid Speech Production

The analysis we initially present sets the stage for some of the others and also leads to an important elaboration of the idea of low-level invariance. Subjects were asked to recite word sequences of different composition and length, specified well in advance of a "go" signal, so as to complete the utterance as soon as possible (Sternberg, Knoll, et al., 1989). The utterances were either simple counting sequences of lengths one through five ("one, two" or "one, two, three, four," for example; *digit condition*), or the same strings with *and* interpolated ("one and two and three," for example; *and-digit condition*) or the same strings with *minus* interpolated ("one minus two minus three," for example; *minus-digit condition*). Within each condition one can regard the strings of increasing length as being generated by starting with *one* and adding elements of a particular "size," the elements being *digit*, *and digit*, and *minus digit*, for the three conditions, respectively, where *digit* represents a digit name and receives a primary stress.

Rather than increasing linearly with the number of elements, n, the durations of utterances produced with this procedure (*sequence durations*) are usually well fitted by a quadratic function:

$$\hat{D}_n = \hat{\alpha} + \hat{\beta}n + \hat{\gamma}n(n - 1).$$ (1)

The nonlinearity reflects the fact that longer sequences are produced at lower rates. We adjust for end effects (such as prepausal lengthening) by forming $\hat{D}_n - \hat{\alpha}$, divide by number of elements, n, and add $\hat{\gamma}$, to arrive at the *mean element duration*:[8]

$$\hat{d}_{.n} = \frac{\hat{D}_n - \hat{\alpha}}{n} + \hat{\gamma} = \hat{\beta} + \hat{\gamma}n.$$ (2)

That is, the mean time between corresponding points of successive elements increases linearly with the number of elements in the utterance (*utterance length*), by an amount given by the parameter γ. Quadratic functions also

---

[8]Duration of an element here includes the durations of all events that occur from its beginning to the beginning of its successor or (for the last element) to the end of the utterance. The addition of $\hat{\gamma}$ is unimportant for our present purposes; it is explained in Sternberg, Knoll, et al. (1989).

account well for the durations of sequences of n distinct keystrokes in typing (Sternberg, Monsell, et al., 1978) where a similar analysis also leads to a linear function of n for mean element duration.[9]

Next we consider the relation between the arbitrary *element*, (such as the word or syllable) in terms of which we design experiments and analyze the data, and the action *unit*, a theoretical entity. We do so here because the result enables us to develop an argument for the existence and identity of an action unit in these rapid utterances: Decomposition of a sequence into units defines a minimal hierarchical structure, and thus permits a test of low-level invariance. Suppose there exists a unit of speech production, an entity that is defined by the scope of an execution process and that is invariant across utterances of different lengths and compositions. We define the *unit period* as the time from the beginning of one unit to the beginning of the next; note that the unit period includes both the time to execute the unit (a low-level process) and the time to invoke its successor (a higher level process). Now consider the effects of increasing the number of equal-size elements in an utterance. Suppose the following equivalence requirement for action units: Regardless of the elements comprising the utterance or of the size of the unit (e.g., the number of phonemes it contains) *mean unit period is either invariant or is increased by the same amount per added unit.* (If unit execution time is itself altered, we have a failure of low-level invariance. If any increase in unit period can be identified with the higher level process, however, then low-level invariance applies to unit execution time itself, and the requirement of a hierarchical control structure is satisfied. The invariance relative to unit size expressed by the equivalence property is then an expression of high-level invariance.) We wish to determine whether there exists such a unit and, if so, to identify it.

Let the entities we designate as the elements of a particular utterance each contain k such units ($k > 0$). Let δ be the mean increase in unit period per added unit. Then the mean increase in unit period per added *element* (i.e., per added k units) is $k\delta$, and the mean increase in *element* duration (i.e., in the summed periods of k units) per added element, or the slope of the element-duration function, is $k(k\delta) = k^2\delta$. (For example, suppose the unit is the syllable and elements are disyllabic words, so that $k = 2$. Let 10 msec be the mean increase in the time from the beginning of one syllable to the next, per added syllable. Because $k^2 = 4$ the mean increase

---

[9]For typewriting (see Fig. 1.11) we use a slightly different parametrization of the quadratic function [$D_n = \alpha + \beta(n - 1) + \gamma n(n - 1)$] and definition of mean element duration [$d._n = (D_n - \alpha) / (n - 1) = \beta + \gamma n$]. These differences are dictated by an interpretation of the duration of an utterance as running approximately from the *beginning* of the first element to the *end* of the last, whereas in those typing experiments where only strokes (taken to be the ends of elements) are recorded, we consider that sequence duration runs approximately from the *end* of the first element to the end of the last.

TABLE 1.1.

Expected and observed slopes (γ) of element-duration functions for three element sizes. Elements are shown in bold type. The slope in the control condition, shown in parentheses, is used to estimate δ.

Condition (msec/element)	Units/Element and Expected Slope						Observed Slope
	Syllable Unit		Word Unit		Stress-Group Unit		
	k	$k^2\delta$	k	$k^2\delta$	k	$k^2\delta$	
1 2 ...	1	$\delta = 8$	1	$\delta = 8$	1	$\delta = 8$	(8.0)
1 and 2 ...	2	$4\delta = 32$	2	$4\delta = 32$	1	$\delta = 8$	6.2
1 minus 2 ...	3	$9\delta = 72$	2	$4\delta = 32$	1	$\delta = 8$	8.7

in the time from one disyllable element to the next is 40 msec per added disyllable.) For utterances composed of different elements, as in the three conditions of the experiment described above, alternative hypotheses about the nature of the action unit lead to alternative sets of k-values (the numbers of units per element), and thus to alternative expectations about relations among slopes of the element-duration functions. By comparing these expectations to the observed slopes we thus develop evidence that bears on the existence of the action unit and its identity.

In Table 1.1 we show the implications for slope ratios of three simple assumptions about the action unit, given that n represents utterance length in elements. If the unit is the *syllable*, then for the *digit*, *and-digit*, and *minus-digit* conditions the numbers of units per element (values of k) are 1, 2, and 3, respectively. If the unit is the *word*, these values are 1, 2, and 2. Finally, if the unit is the *stress group* (composed of a stressed syllable and any associated unstressed syllables), then the values are 1, 1, and 1. If we analyze the data in terms of number of elements, then the predicted ratios of slopes of the element-duration functions generated by our three definitions of the unit are 1:4:9 (syllable unit), 1:4:4 (word unit) and 1:1:1 (stress-group unit). The observed values (Table 1.1; functions themselves in Fig. 1.2) clearly favor the stress-group unit.[10] That these values are consistent with expectations based on any one of these simple assumptions supports the existence of an action unit, as defined above. A stress-group unit is also consistent with the finding of parallel element duration functions for utterances composed of unrelated monosyllabic and disyllabic nouns (Sternberg, Monsell, et al., 1978) and for utterances composed of monosyllabic words and phonologically matched nonwords (Sternberg, Wright, et al., 1980). Approximate equality of slopes of the element-duration functions (Equation 2) when n denotes the number of units implies approximate

[10]Thus the number n of units happens to equal the number of those elements in terms of which (in this case) we elected to manipulate the length of the utterance. For further discussion of the stress group as a unit see Fowler (1981, 1985) and Sternberg, Knoll, et al. (1989).
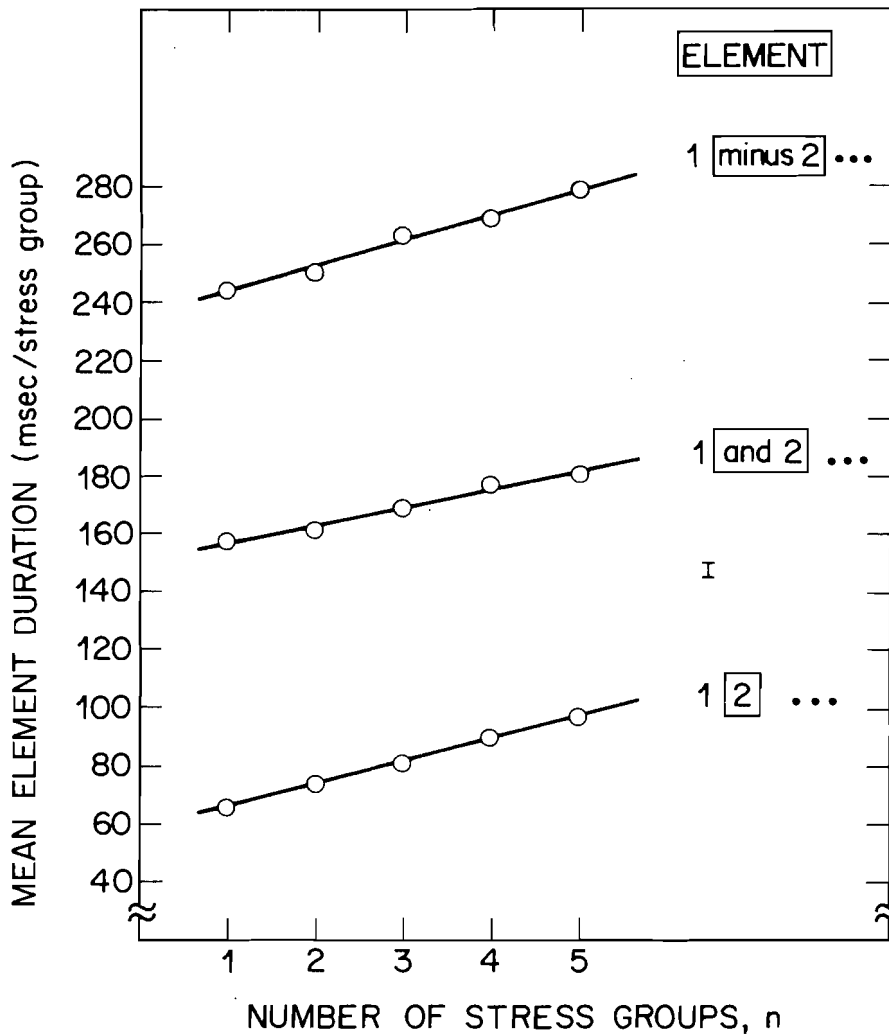
FIG. 1.2. Mean element-duration functions based on stress-group elements for control utterances (*digit* condition) and for utterances with interpolated unstressed words (*and* or *minus*), with estimate of ± SE and fitted linear functions: 58 + 8.0n (*digit* condition); 151 + 6.2n (*and-digit* condition); 235 + 8.7n (*minus-digit* condition). Results are averaged over four subjects and three days; about 140 (70) observations per point for control (other) conditions.

equality of the nonlinear part of the sequence-duration function (Equation 1). Sequence-duration functions for utterances containing units of different size can then differ by at most a linear function. We see later how this constraint follows from our two invariance properties.

For our present purposes the main point is that we have a set of data that is accounted for very nicely by a particular definition of the action unit. Having identified a unit, we now ask whether that unit functions as a constituent in a hierarchical structure. We first consider whether it obeys low-level invariance; later we discuss a test of high-level invariance, as units of different size are exchanged.

Further Analysis by Segmentation of Single Speech Units

For the test of low-level invariance, the property of the unit that we consider is its mean duration; the influencer is the number n of units in the utterance. It is clear that if we identify unit duration with unit period, as plotted in Fig. 1.2, we have a violation of low-level invariance. The three approximately parallel functions indicate two effects on unit period, effects that combine additively. First, the vertical displacement of the three functions shows, unsurprisingly, that the *size* of the unit—the number of phonemes or syllables it contains—has a substantial effect on its period. That this effect is independent of n suggests the possibility that unit execution time, a component of the unit period, might be invariant relative to sequence length. Second, the fact that instead of being flat, the element-duration functions rise systematically, shows that the *number* of units in the utterance also has an effect on unit period. Here we have what appears initially to be a clear violation of low-level invariance: a characteristic of a purported hierarchical constituent depends on its context. However, the analysis of how the effect of utterance length on unit period is distributed within that period provides a means by which low-level invariance can be preserved.

In two experiments conducted under conditions similar to those already described (Sternberg, Wright, et al., 1980; Sternberg, Knoll, et al., 1989), utterances were composed of words selected such that segmentation based on the acoustic signal would be feasible: Words were initially stressed disyllables (e.g., *copper, token*) in which each syllable starts with an articulatory closure. Each word in the second experiment (1989) was decomposed into 6 segments, corresponding roughly, for each syllable, to the closure, release, and vowel: $C1$, $R1$, $V1$, $C2$, $R2$, $V2$. The upper panel of Fig. 1.3 displays the element-duration function from that experiment, comparable to the functions shown in Fig. 1.2, here with a slope of 11 msec/word. The lower panel shows the segment-duration functions, which indicate how the effect of the number of units in the utterance is distributed within the stress-group unit.[11] Rather than pervading the unit, the effect of utterance length is almost entirely restricted to $V2$, the vowel of the second syllable; durations of the 5 other segments are virtually invariant. It is possible that the second vowel is prolonged until higher level control structures have invoked the subprogram for the subsequent unit. Low-level invariance is required of the time to execute a subprogram, not of the time to invoke it. We thus have what appears to be a qualified form of low-

---

[11]The segment-duration function for segment i (here, $1 \le i \le 6$) is obtained by starting, for each n, with the summed durations of the n ith segments in utterances of length n, and using the resultant set $\{D_{ni}\}$ to determine a set of n mean segment durations $\{d._{ni}\}$ for the ith segment, just as utterance durations $\{D_n\}$ are used to determine mean element durations $\{d._n\}$. The observed and fitted element-duration functions are then sums of the observed and fitted segment-duration functions, respectively.
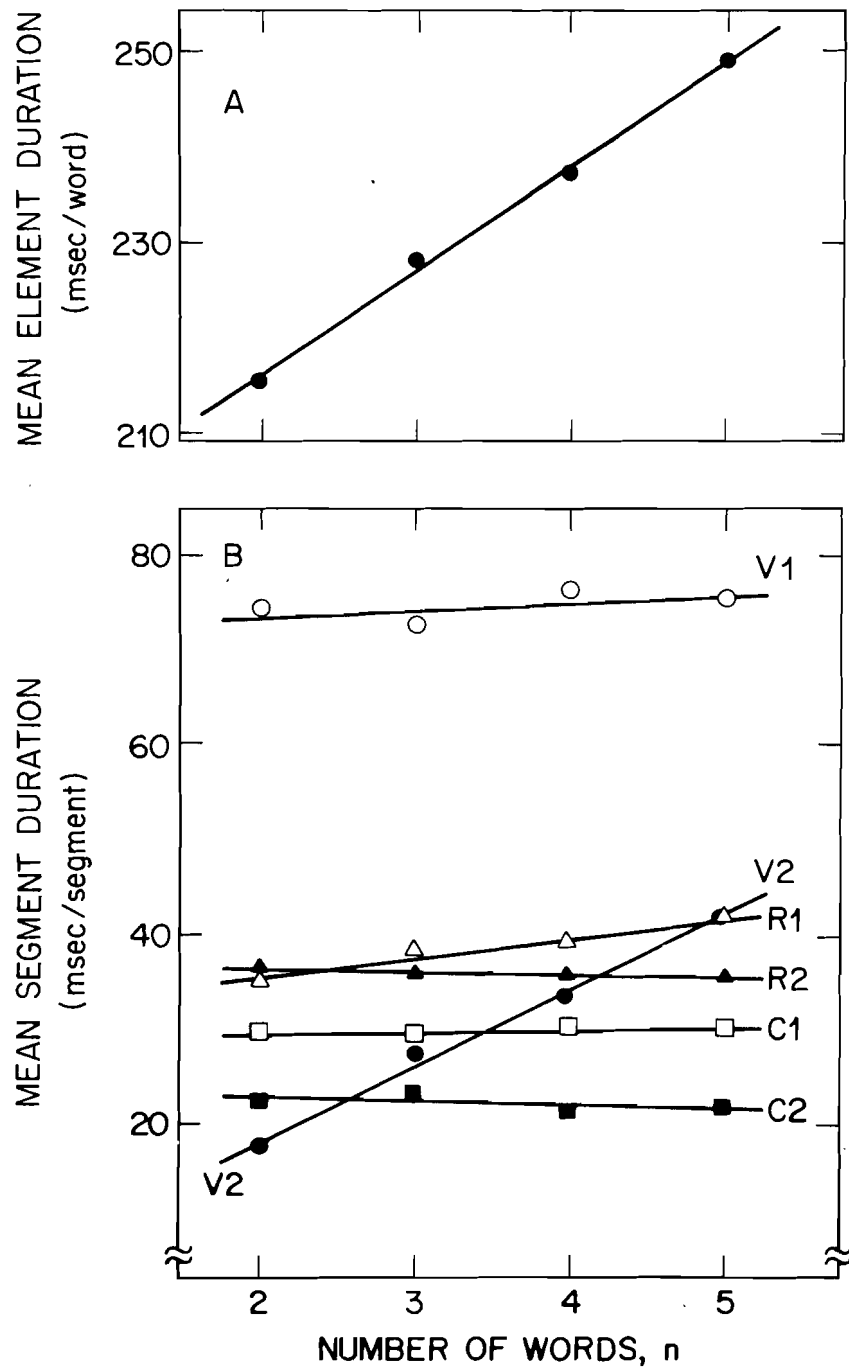
FIG. 1.3. Mean element duration (with fitted function $194 + 11n$ msec) and mean segment durations of two-syllable spoken words. Results are averaged over three subjects and two days; about 100 (50) observations per point for lists of $n > 2$ ($n = 2$) words.

level invariance which, on reflection, is what we might have expected: Given that speech is relatively continuous, that the time required to invoke a subprogram may vary with sequence length (because length influences the control structure), and that smoothing occurs to provide for continuity, not all speech segments would be expected to demonstrate duration invariance. That the effect of utterance length is seen in one epoch per stress group (indicating one invoking process per stress group) is consistent with the conclusion about unit size based on properties of the mean element-duration function. The unit period thus appears to include two additive components, one attributable to the unit and dependent on unit size alone, and the other attributable to higher level control operations and dependent on number of units.

A difficulty remains, however. Although we find approximate duration invariance relative to sequence length for 5 of the 6 segments of the stress-group unit when the mean is taken over serial position, low-level invariance also requires that these segment durations be invariant over serial position; and this is not uniformly the case: instead, serial position influences the duration of $V1$, the first syllable's vowel, which tends to increase with serial position (Sternberg, Wright, et al., 1980). One possible way to rescue the invariance property would be to argue that prosodic effects—including effects of serial position on timing—are introduced *after* the hierarchical control structure has done its work, so as to modulate its output.[12] Low-level invariance would then characterize a hypothetical representation that antedates the action sequence itself. With this elaboration of the hierarchical control model we turn our attention away from speech production and toward sequences of finger movements.

Tests in Keystroke Sequences Containing Repeated Pairs of Distinct Strokes

Evidence of hierarchical control in strings of eight finger taps.   Rosenbaum, et al. (1983) have provided what appears to be persuasive evidence of hierarchical control in the production of patterned strings of finger taps. They studied the performance of strings of eight taps made by the fingers on alternating hands and containing repeated pairs. An example is R_1, L_1, R_1, L_1, R_2, L_2, R_2, L_2, where R and L denote left hand and right hand, and subscripts 1 and 2 denote index finger and middle finger, respectively. Subjects were asked on each trial to produce one of a set of such strings repeatedly without interruption and as fast as possible, for six renditions.

---

[12]The idea that the motor system temporally modulates or "interprets" the output of a higher level control system by introducing delays that might have some structure has been suggested for isochronic taps by Wing (1977).

Evidence for hierarchical control was derived mainly from this pattern of mean intervals between successive taps: The mean intertap intervals (*ITIs*) defined by the four pairs of taps 1-2, 3-4, 5-6, and 7-8 (*first set*) were similar and shortest; they were longer and similar for the two intervals defined by taps 2-3 and 6-7 (*second set*); and they were longest for the two intervals defined by pairs 8-1 and 4-5 (*third set*). Furthermore, the increases in mean *ITI* from first set to second and from second set to third were approximately equal, consistent with a recursive process.

The control structure suggested by Rosenbaum et al. is an ordered binary tree (Fig. 1.4, $n = 8$) a special case (restricted to binary) of the phrase marker discussed earlier. Thus, each of the two nodes below the first bifurcation in the tree representation of the eight-tap sequence specifies a four-tap constituent, each of the four nodes below the next bifurcations specifies a two-tap constituent, and each of the eight final, bottom-level nodes specifies a single tap. The tree structure is augmented, for control purposes, by a ("preorder") tree-traversal process (Aho, Hopcroft, & Ullman, 1983) that starts at the top (root) node and proceeds to the lowest level nodes from left to right, in turn. To repeat the sequence the process passes from rightmost node to root, and then proceeds as before. The intervals between successive taps are assumed to increase linearly with the number of nodes that must be traversed; the resulting model fits the mean *ITI* data reasonably well. (In an isomorphic conceptualization of such a tree and its control process, each node corresponds to a subprogram of the program associated with its parent node, and tree traversal is replaced by a "subprogram-retrieval process" with properties as in Sternberg, Monsell, et al., 1978.)

Tree traversal is one control process associated with a hierarchical structure that operates by transfer of control and requires low-level and high-level invariance of the sequences it produces. Thus (low-level invariance) once control is passed to a lower level node the traversal process for that node is independent of the structures connected to it through higher level nodes (from which control was derived earlier and to which it is transferred later); similarly (high-level invariance) the traversal processes in those other higher and disjoint parts of the structure are independent of what the process does in relation to the lower level node. Although strings of taps known to have the suggestive pattern of mean *ITIs* described above form a very limited set, it is worth testing the invariance properties in such strings, partly because hierarchical control is strongly indicated by the *ITI* data, and partly because the model advocated by Rosenbaum et al. (1983) requires these properties.

Modifications and elaborations in a keystroke experiment. We therefore proceeded to test low-level invariance in strings with repeated pairs, by modifying and elaborating the experiment of Rosenbaum et al. (1983).
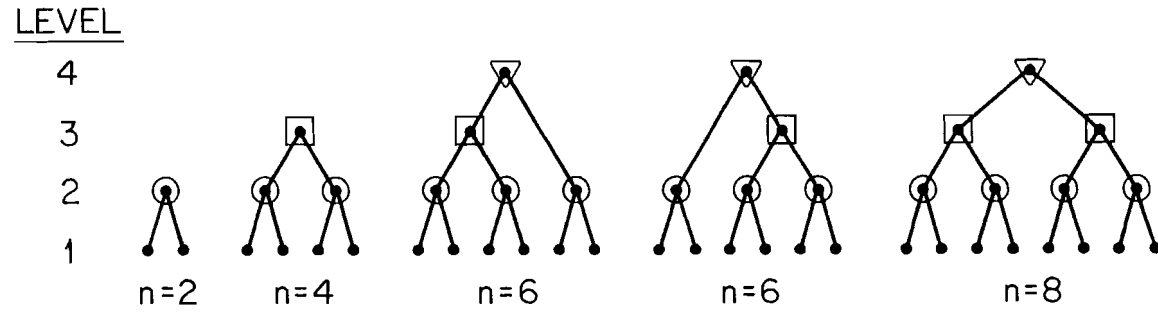
FIG. 1.4.   Binary trees for strings of 2, 4, 6, and 8 letters. Nodes at second, third, and fourth control levels are designated by circles, squares, and triangles, respectively.

TABLE 1.2.

Repeating-subsequence letter strings of four lengths. The letters *j*, *f*, *k*, and *d* are typed with the right index finger, left index finger, right middle finger, and left middle finger, respectively.

Repeating-Subsequence Strings	
$n = 2$ jf (. . .)	kd (. . .)
$n = 4$ jfkd (. . .)	kdjf (. . .)
$n = 6$ jfkdkd (. . .)	kdjfjf (. . .)
$n = 8$ jfjfkdkd (. . .)	kdkdjfjf (. . .)

The principal differences were as follows: First, we studied typed rather than tapped sequences. One reason was to more closely approximate the procedure we had used in a series of studies of short keystroke bursts deliberately designed to contain no repeats (Sternberg, Monsell, et al., 1978). Second, in addition to strings containing obvious substrings of length four (*pair-quad strings*), we also used strings containing no such substrings; an example of such a *pair-only string*, using the notation above, is R_1, L_1, R_1, L_1, R_1, L_1, R_1, L_2. Third, in addition to a *recycling* condition in which each string was performed six times with no break, we also used a *one-cycle* condition with the pair-quad strings, in which subjects had to produce only a single rendition of the string on each trial. (Although this variation had an effect, it had no important bearing on the hypotheses under test.) Finally, we used sequence length and serial position as influencers in tests of low-level invariance.[13]

Examples of the strings we used are shown in Table 1.2. Those of length $n = 8$ required finger patterns similar to those used by Rosenbaum et al.; the shorter ones ($n = 2, 4, 6$) were substrings selected to have no period smaller than n.[14]

---

[13]Although expected from several considerations, it had not been known that sequence length influences mean performance rate in such strings. For sequence length to be determinate under recycling conditions, a string of length n must have no period smaller than n. Thus, whereas *jfkdkd* or *jfjfjd* are acceptable length-6 strings, *kdkdkd* is not: 6 renditions of the latter string is equivalent to 18 renditions of a length-2 string.

[14]In the first part (A) of the experiment (4 practiced subjects) we examined only recycling performance of the 8 strings in Table 1.2. Subjects practiced a string ad lib before initiating the trial, and were signaled after 6 renditions, with no penalty for overrunning. Data were taken in 4 brief sessions of 48 trials per subject. In the second part (B) of the experiment (3 practiced subjects) we examined the same strings in both recycling and one-cycle conditions, and also studied pair-only strings of 4 lengths in the recycling condition, with similar procedures. One-cycle data were taken in sessions both before and after the recycling data for pair-quad strings. Data from pair-only strings (*jf, jfjd, jfjfjd, jfjfjfjd* and *kd, . . ., kdkdkdkf*) were taken in later sessions.
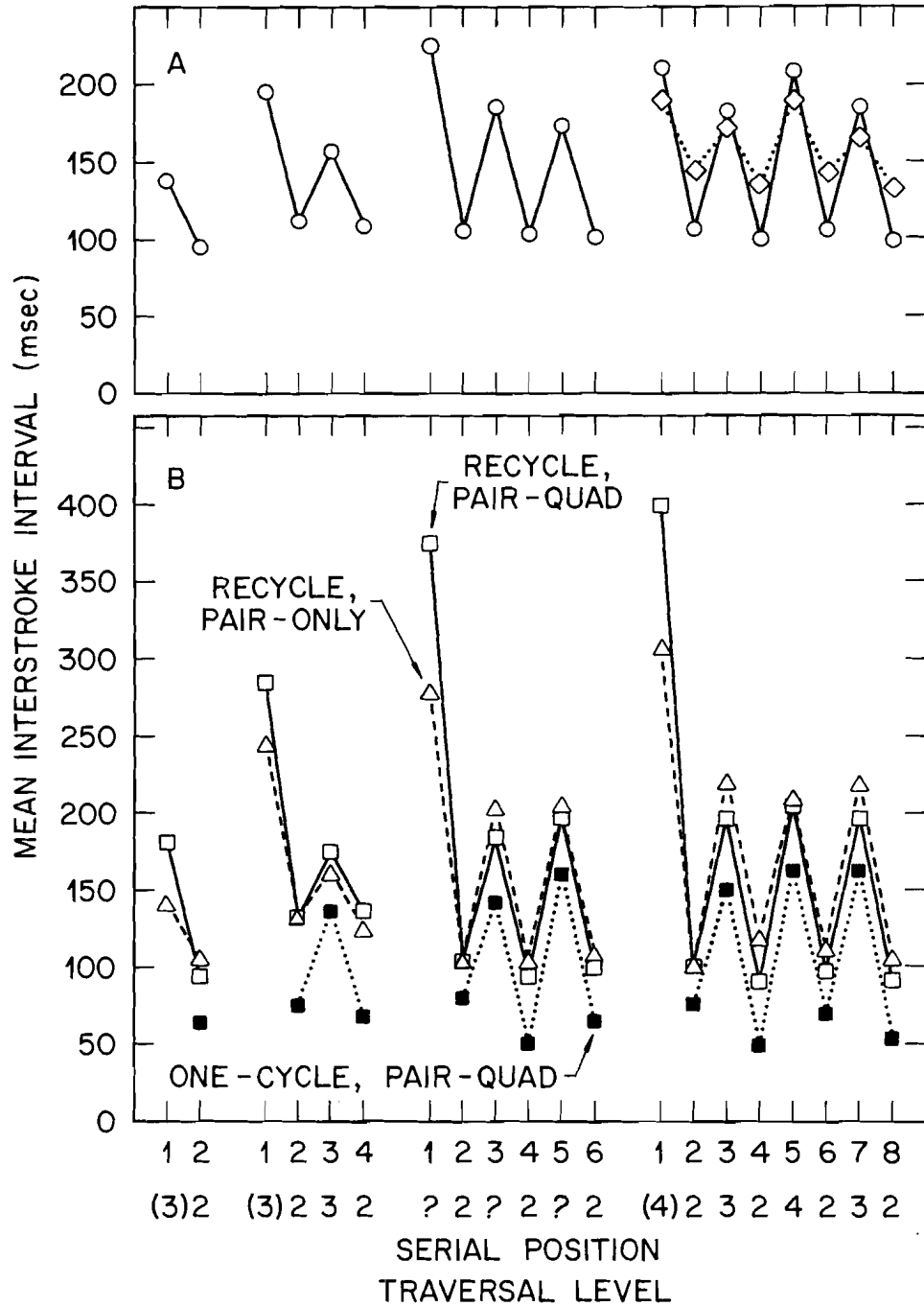
FIG. 1.5.   Mean Interstroke Interval (*ISI*) as a function of serial position for strings of 2, 4, 6, and 8 letters. The serial position of an *ISI* denotes the stroke with which it terminates. Numbers in the bottom row indicate *traversal level*: the level of the highest node traversed in generating each *ISI*, according to the binary tree-traversal model. Traversal levels in parentheses depend on an auxiliary assumption in the model. The time interval from the "ready" signal to the initial stroke has been excluded.

A: Mean *ISI* from the four subjects in Part A of the experiment, with 48

Results of the modified experiment.   Before considering tests of low-level invariance we comment on some other aspects of the data. Mean *ISI* as a function of serial position in strings of four lengths is shown in Figs. 1.5A and 1.5B. Also shown, for comparison, are mean *ITI*s in the study of Rosenbaum et al. (1983), for pair-quad strings with $n = 8$. Qualitatively our findings for $n = 8$ strings agree with those of Rosenbaum et al.: First-set *ISI*s (see definition above; second-level traversal) are the shortest, and third-set *ISI*s (fourth-level traversal) are the longest; magnitudes of the differences are greater and the effect of assumed traversal level less linear in the typing than the tapping data.

One difficulty presented by these data for the model of Rosenbaum et al. is the systematically large value of the first *ISI* under recycling conditions. Although not required by the spirit of the model, it attributes the time between the end of one rendition and the beginning of the next to traversal of the tree from the rightmost to the leftmost node. For $n = 2$, for instance, this traversal covers the same distance as the one presumably required between the first and second stroke, so the marked serial-position effect is unexpected. Similarly, the large difference (especially for subjects in Part B) for strings of length $n = 4$ between the first and third *ISI*, which also should involve equal traversal distances, is unexpected. Structural ambiguity for $n = 6$ precludes this comparison; the corresponding effect for $n = 8$ (difference between first and fifth *ISI*) is negligible in Part A but large in Part B. Because there appears to be a systematic inflation of the first *ISI*, perhaps attributable to a mechanism that counts renditions, we exclude it from further tests.

Two tests of low-level invariance.   Given this experiment, there are two different tests of low-level invariance that we might perform, of differing generality. In the *binary-hierarchy test* (the less general), we assume that if the control structure is hierarchical then it must be binary, as proposed by Rosenbaum et al. (1983). The resulting structures for strings of four lengths are shown in Fig. 1.4. The figure shows that longer strings are distinguished by having higher level nodes; indeed, any slowing of performance with increasing length—any increase in mean *ISI*—is due entirely to the increasing depth of the tree, that is, to the augmentation of

---

trials per sequence length per subject; about 600 observations per point. Also shown, for $n = 8$ are the tapping data from Rosenbaum et al. (1983), from five subjects; about 6000 observations per point.
B: Mean *ISI* as a function of serial position from the three subjects in Part B of the experiment, under three conditions: Recycling pair-quad strings (as in Part A; about 850 observations per point); One-cycle pair-quad strings (about 500 observations per point); Recycling pair-only strings (about 850 observations per point).

the set of traversals required to execute the sequence by traversals of increasing length (and duration) as well as additional traversals of the same lengths. As we move from $n = 2$ to $n = 4$, for example, we add, within the string, an additional traversal of a node at the second level (associated with the second pair), as well as a traversal of a third-level node (between the two pairs). The structure is ambiguous for $n = 6$; whereas we use $ISIs$ from hypothesized second-level nodes in $n = 6$ strings to test second-level invariance, we therefore limit our tests of third-level invariance to $ISIs$ generated by hypothesized traversals of third-level nodes in strings of lengths $n = 4$ and $n = 8$.

In a *general-hierarchy test* we do not constrain the control structure to be binary, or the control process to be tree-traversal. As we vary an influencer such as sequence length in a test of low-level invariance, however, we cannot permit the hierarchical structure to be completely free. Recall that the property in question is the invariance across changes in the control structure of the process that executes a constituent substring. Testing this property requires that the *same* constituent exist under different levels of the influencer. Given a binary-tree control structure, this property follows *a fortiori*: the lowest level constituents of the hierarchy are *pairs* of elements for strings of all (even) lengths. However, if we impose no constraints on the structure, then the lowest level constituents might change as the influencer is varied. Thus, the execution of strings of several lengths might indeed by hierarchical, but the length difference might correspond to differences in the size of the lowest level constituents, or to differences in the size of constituents at all levels, rather than to differences in the depth (extent of bracket nesting) of the structure. In that case if we assumed fixed low-level constituents we might conclude against low-level invariance for the wrong reasons. Especially for strings constructed from repeated pairs as in the present experiment, but even for strings with no repeats, such flexibility in the organization of low-level constituents is implausible, whereas increasing the size of the structure by increasing its depth is in the spirit of the recursiveness of hierarchical control.

In connection with the present experiment we restrict ourselves to the binary-hierarchy test. Insofar as this test succeeds, the constraint is unimportant; insofar as this test fails, one from the more general class might succeed.

Application of the binary-hierarchy test.   For $ISIs$ after the first in pair-quad strings we note that although Part B reveals large and significant differences between one-cycle and recycling conditions ($p < .05$ for $n = 2, 6, 8$ strings, despite the small number of subjects), the effects of serial position at each length ($n = 4, 6, 8$) are statistically indistinguishable

(interactions with one-cycle versus recycling not significant); we therefore examine these effects for the two conditions combined. Considering all positions after the first, Parts A and B both show large and significant effects of serial position, inviting us to test low-level invariance relative to this influencer.

Combined with the binary-hierarchy model, low-level invariance relative to serial position requires that for each string length, mean *ISIs* for second traversal-level (first-set) stroke pairs are equal, regardless of serial position. Instead there is systematic perturbation in all pair-quad strings, in which the *ISI* tends to be reduced by immediate repetition of a pair, which reaches significance in Part B ($p = .05$ for $n = 6, 8$). This effect is small relative to the large range of variation of *ISI* with serial position, so should perhaps not be given great weight; it reminds us of the difficulties for low-level invariance presented by prosodic (serial position) effects in speech. Low-level invariance relative to serial position also requires equality of the two third-level *ISIs* in $n = 8$ strings (indexed by 3 and 7 in Fig. 1.5), a requirement that is nicely satisfied.

We test low-level invariance with respect to sequence length by examining *ISIs* associated with second and third traversal levels. Before considering the effect of sequence length on these *ISIs*, we note that the overall mean *ISI* increases from $n = 2$ to $n = 4$, and then rises only slightly.[15] Sequence length does not therefore qualify as an influencer from $n = 4$ to $n = 8$; low-level invariance observed over that range would not be very persuasive. Means over the *ISI* values associated with purported second-, third-, and fourth-level traversals are shown in Fig. 1.6 as a function of sequence length, separated by condition and part of experiment. We see a systematic effect of sequence length at the second level, in the form of a peak at $n = 4$; the effect is significant in both Parts A ($p = .04$) and B ($p = .03$). At the third level we also appear to have a failure of invariance, with the mean *ISI* increasing with sequence length.[16]

We thus have three pieces of evidence against low-level invariance in the performance of patterned keystroke strings: Effects of serial position on first-set *ISIs*, and of sequence length on first-set and second-set *ISIs*. Even if we use these findings to reject the binary model, however, we

---

[15]Mean *ISIs* in msec over $n - 1$ serial positions (excluding *ISIs*, in recycling conditions, that terminate with the first stroke) for pair-quad strings of lengths $n = 2, 4, 6,$ and 8, respectively, are (Part A, recycling) 95, 126, 134, 141; (Part B, recycling) 94, 148, 136, 139; and (Part B, one cycle) 64, 94, 100, 104. For pair-only strings (Part B, recycling) they are 106, 140, 146, 156.

[16]The increase is shown by 6 of the 7 subjects ($p = .06$ by a one-tailed binomial test), but there is wide variation in the size of the effect, which thus needs further substantiation.
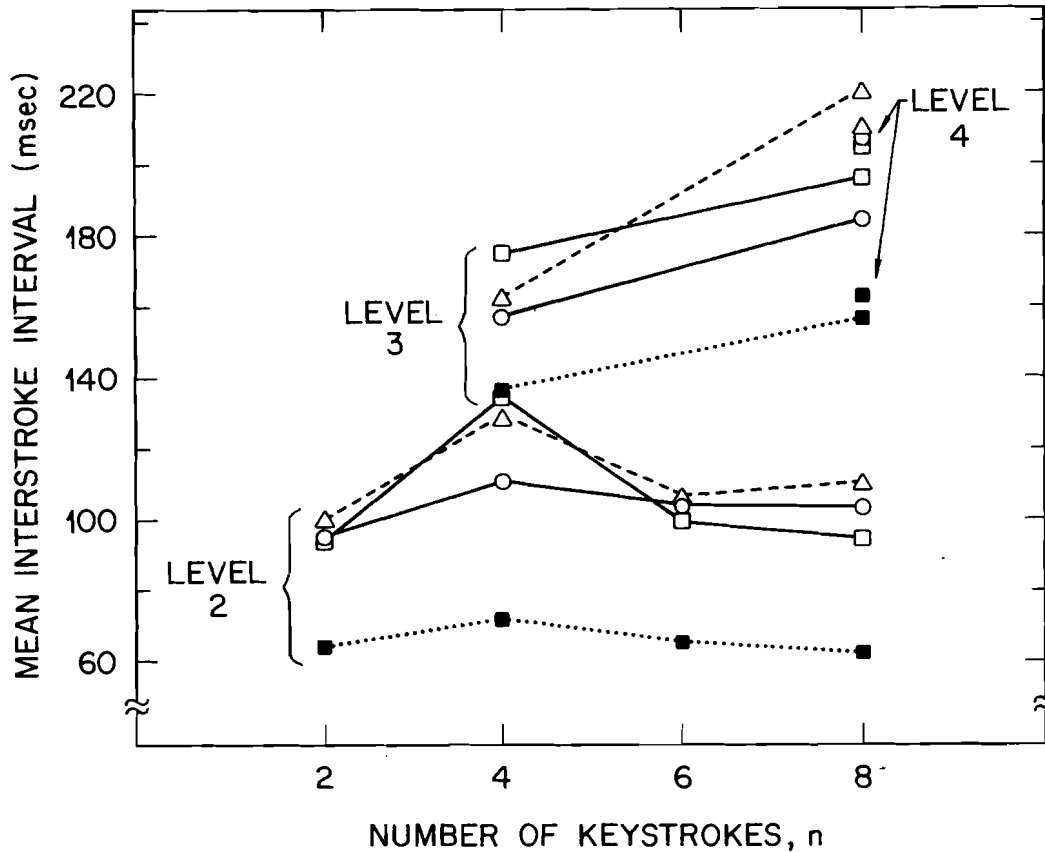
FIG. 1.6. Mean *ISI* as a function of sequence length and traversal level (see Fig. 1.5). The four conditions are designated as follows. From Part A: pair-quad, recycling, open circles. From Part B: pair-quad, recycling, open squares; pair-quad, one-cycle, filled squares; pair-only, recycling, open triangles.

might be able to avoid rejecting the more general hierarchical model. For example, the failure at the third level, relative to sequence length, could be explained if the constituents at that level for $n = 8$ contained one pair and three pairs rather than two and two, and if the number of pairs controlled by a third-level node had a disproportionate effect on the associated *ISI*.

Test of Constituent Invariance in Strings of Distinct Strokes

Assumptions for restricted and general tests; expected linkage proportions. In the case of patterned strings such as those examined above there are plausible mappings of the string onto a hierarchical control

structure. For a string of distinct strokes (with no repeats), however, there is little basis for specifying the mapping between string and structure. For such strings we would thus like to test the hypothesis that some stroke pairs are linked together in constituent substrings (at the second level), but without specifying which strokes are contained in these constituents. Our test in this case is therefore applied to the distribution of *ISI*s pooled over all pairs within the string; we ask about the proportion of the adjacent stroke pairs in the string that are linked together in low-level substrings.

Again, we consider both a restricted and a more general test. In the restricted test we assume, with Rosenbaum et al. (see Footnote 20) that the hierarchy is binary—even for strings with no repeats. For even-length strings the lowest level substrings are thus composed exclusively of pairs; for odd-length strings they contain pairs plus a single one-element constituent (in an unknown position). For this restricted class of hierarchical structures we can thus specify the number of pairs of adjacent strokes that must be linked together in low-level (two-stroke) substrings: In a string containing either $2k$ strokes ($2k-1$ possible pairings; $k \geq 1$) or $2k+1$ strokes ($2k$ possible pairings), a binary hierarchy will contain k actual linked pairs at its lowest level. The linkage proportions are thus $k/(2k-1) > 0.5$ and $k/2k = 0.5$, respectively, or at least 0.5 regardless of string length.

As mentioned earlier, any general test of low-level invariance must jointly test the assumption that identities of low-level constituents are fixed over levels of the influencer. The specific form of the assumption that we require here is weak, however: We assume that at least one of the lowest level substrings of size $s \geq 2$ (containing $s-1$ linked pairs) is of the same size in long and short strings. Again, we justify this assumption by the recursiveness naturally associated with the idea of hierarchy: Changes in the size of a structure are thus expressed by changes in its depth rather than in the sizes of its low-level constituents. Given the assumption, the linkage proportion for a string of length n ($n-1$ possible pairings) is no less than $(s-1)/(n-1)$. Because $s \geq 2$, the linkage proportion is thus bounded below by $1/(n-1)$.

In summary, the linkage proportion for a binary hierarchy can be no less than 0.5; for a general hierarchy and for strings of no more than five strokes, it can be no less than 0.25.

Our tests require one additional assumption: The interstroke times of pairs linked at the lowest (second) level tend to be shorter than other interstroke times. This is obviously true for a tree-traversal model and is plausible for any hierarchy model.
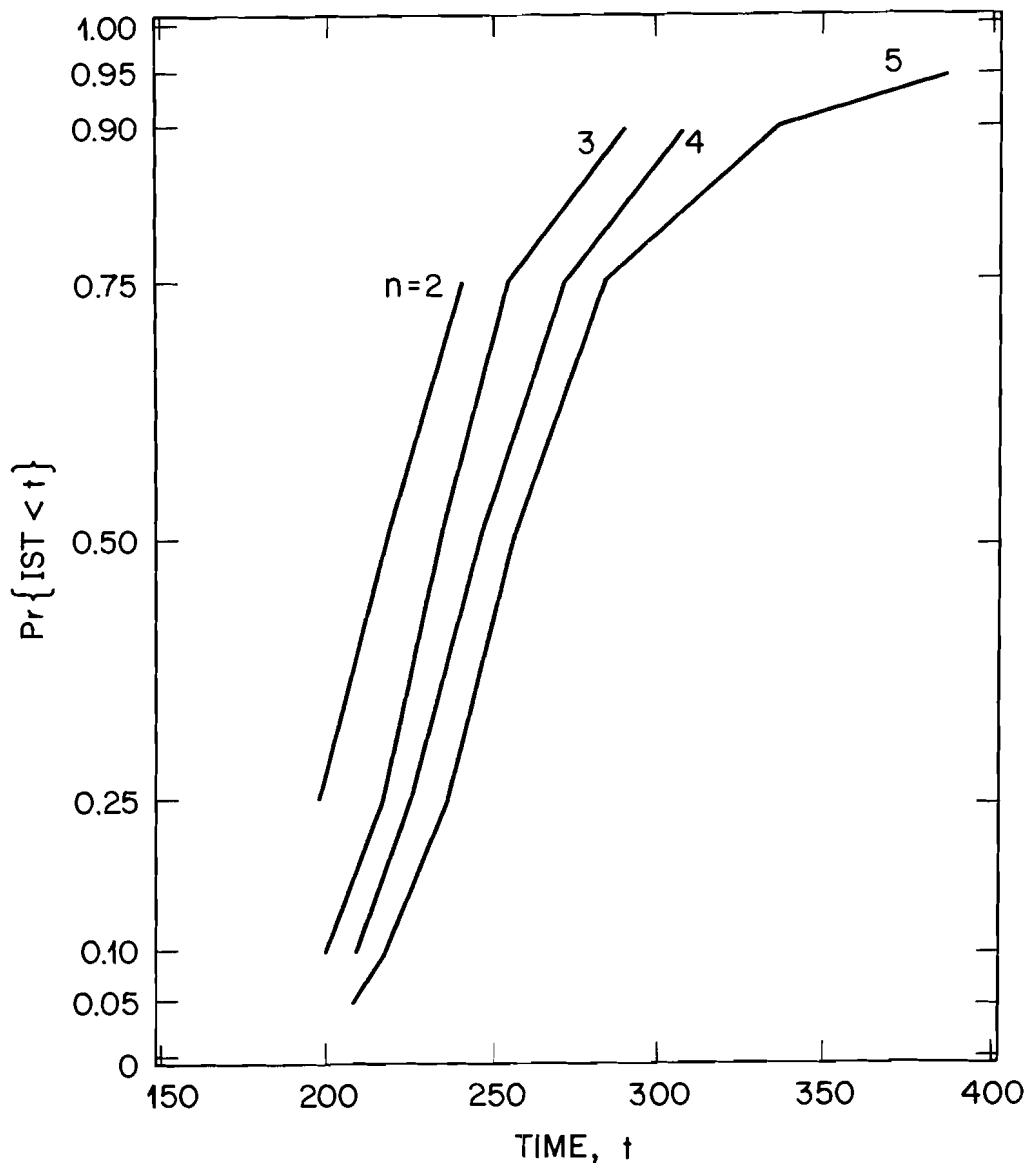
FIG. 1.7.   Mean interstroke-time distributions for sequences of length
$2 \leq n \leq 5$. See Footnote 17 for details of estimation.

Observed linkage proportion: Upper bound. The test is based on the following idea: If strings of two lengths (for example, $n = 2$ and $n = 5$ with *ISI* distributions H_2 and H_5 respectively) contain invariant substrings, then the shortest *ISI*s in H_5 should be as short as the shortest *ISI*s in H_2. The two cumulative distributions should therefore diverge or fan out from the same low-probability point. What we observe, however, are distribu-

tions better described as being displaced on the time axis, or translated, even at low probability values, which suggests that the shortest ISIs in H_5 are longer than the shortest in H_2. Fig. 1.7 shows mean ISI distributions for sequences of four different lengths ($n = 2, 3, 4, 5$) in an experiment in which subjects typed strings of distinct consonants.[17] In can be seen that these distributions are better described as being displaced than as diverging from a common point.

Without additional assumptions we wish to quantify the extent to which an invariant substring (and hence a hierarchical structure) is precluded by such findings. One way to do so is to express the pooled distribution of ISIs for a string of length n as a mixture of two distributions. One component of the mixture is the distribution of ISIs associated with invariant substrings; this distribution, which we call $G_I(t)$, must be independent of n; its proportion in the mixture (its mixing probability p_n) is the same as the linkage proportion. The other component of the mixture is the distribution of ISIs for which invariance is permitted to fail and which can therefore vary with n; we call it $G_n(t)$; its mixing probability is $1 - p_n$. Our strategy is to determine what the relations among observed distributions can tell us about p_n. The observed distribution of ISIs pooled over serial positions corresponds to the mixture distribution, $H_n(t)$, which may be written in terms of its components as follows:

$$H_n(t) = p_n G_I(t) + (1 - p_n) G_n(t). \tag{3}$$

The assumption that linked pairs tend to have shorter ISIs than unlinked pairs is expressed by a dominance relation: $G_I(t)$ dominates $G_n(t)$. That is, for any t, $G_I(t) \geq G_n(t)$.

Given the invariance of $G_I(t)$ with respect to sequence length n, and the dominance assumption, it is easy to show that for any two sequence

---

[17]The data were drawn from a subset of the trials in a large experiment conducted for different purposes. They are the interstroke times of 6 bigrams, each of which appeared in all possible ($n - 1$) serial positions in strings of length $n = 2, 3, 4$, and 5. Each string was typed by the fingers of one hand; the two hands were used equally often. We first pooled the data for a bigram across serial positions. Next we pooled the pairs of distributions for the two fastest bigrams (*jm, jh*), the two medium-speed bigrams (*fg, hm*), and the two slowest bigrams (*fb, rb*). This procedure provided sample sizes of $6(n - 1)$ for each bigram pair and each subject. We then used linear interpolation to estimate quantiles for a specified set of probability values ($p = 0.25, 0.50, 0.75$ for $n = 2$, these plus $p = 0.10, 0.90$ for $n = 3, 4$, and these plus $p = 0.05, 0.95$ for $n = 5$). Means of the resulting quantiles were then obtained over bigram pairs and subjects to provide the plotted values.

lengths, r and s (e.g., $r = 2$ and $s = 5$), and for all values of t,

$$p_s \leq \frac{H_s(t)}{H_r(t)},$$ (4)

and hence that

$$p_s \leq \min_t \left[\frac{H_s(t)}{H_r(t)} \right].$$ (5)

With only these minimal and plausible assumptions the relation between pooled *ISI* distributions for strings of different lengths can therefore provide us with an upper bound on the mixing probability, and thus on the linkage proportion.[18] It follows that $p_5 \leq \min[H_5(t)/H_2(t)]$, where the minimum is taken over all values of t. For example, if there is any (small) *ISI*, t_0, such that H_2 contains some *ISI*s that small $[H_2(t_0) > 0]$ but H_5 contains none that small $[H_5(t_0) = 0]$, then the H_5/H_2 ratio is zero for $t = t_0$, which implies that p_5 must be zero. Of course, this discussion concerns population distributions, and we must work with samples, which creates an estimation problem; we chose the estimation method so as to be conservative, in the sense of increasing the generosity of the upper bound.[19]

Upper bounds on p_5 estimated for each of eight subjects are represented by the filled circles in Fig. 1.8; The mean such bound (0.21) and $\pm SE$ are also shown. (Open circles are derived from a condition to be described later for which other data suggest a high linkage proportion and a hierarchical structure.) We checked the method by applying it to simulated data from mixtures of several known distributions with specified linkage proportions, and found that the upper bound was substantially greater than the actual linkage proportion in all cases, suggesting that it is a generous bound. For example, an estimated upper bound of 0.2 corresponded to an actual linkage proportion of 0.04 or less.

Even without taking the generosity of the bound into account, these findings clearly permit us to reject the restricted (binary) hierarchy model, which, as we have said, requires a linkage proportion for $n = 5$ of 0.5.

---

[18]To prove the inequality, first set $n = s$ and note that $H_s(t) \geq p_s G_r(t)$ for all t. (The mixture dominates one of its weighted components.) Now set $n = r$ and note that $G_r(t) \geq H_r(t)$ for all t. [Because the invariant distribution dominates the non-invariant, it also dominates any mixture of itself and the non-invariant, and therefore dominates $H_r(t)$.] Now, combine the two inequalities so as to eliminate $G_r(t)$: $H_s(t) \geq p_s H_r(t)$, for all t. The desired inequality follows.

[19]A discussion of estimation methods for essentially the same problem is provided by Sternberg and Knoll (1986).
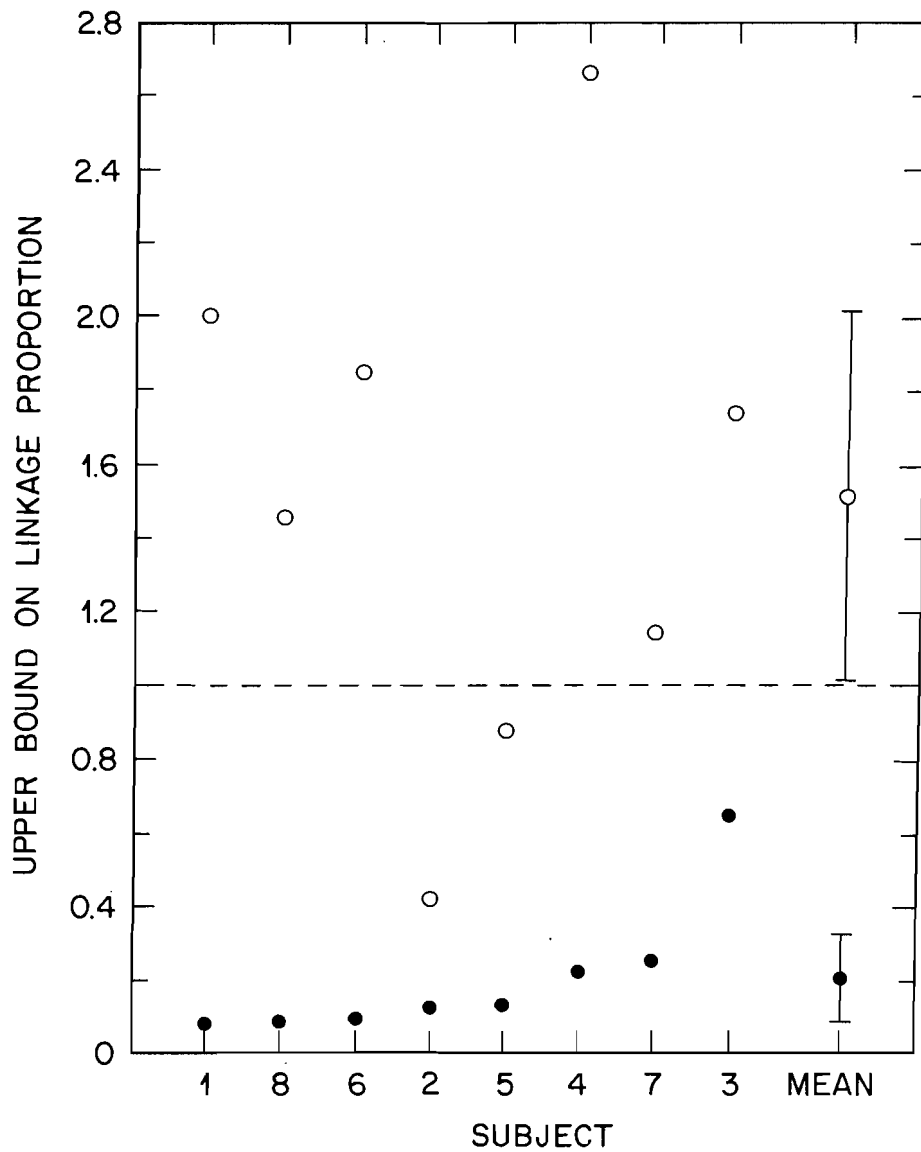
FIG. 1.8. Estimates for eight subjects of the upper bounds for the linkage proportion in keystroke strings of length $n = 5$. Filled circles represent values for strings of distinct strokes. Open circles represent estimates for the same subjects of upper bounds on the linkage proportion for the two strokes of doublets embedded in strings of length $n = 5$. Also shown in each case is the mean and $\pm SE$.

Claims that a traversal process operating on a binary tree can describe the execution of strings of distinct keystrokes thus appear to be incorrect.[20] The present analysis is limited to strings typed by the fingers of the same hand; *ISI* distributions for alternating-hand strings suggest a similar conclusion, however. For the more general model, whose linkage proportion must be at least 0.25, the obtained values of the bound as such do not permit rejection. However, given the generosity of the bound, that model also becomes questionable.

If we turn from strings of distinct strokes to strings containing immediate repeats, or *doublets*, (strings such as *sccrz*) and examine the distribution of *ISIs* for the doublet (the time interval between the first and second *c* in *sccrz*, for instance) in the same way, the estimated lower bound on the linkage proportion (values shown by the open circles in Fig. 1.8) is consistent with the doublet forming an invariant substring.[21] We further consider the case of embedded doublets in a later section. Two implications are relevant to the present argument: First, the finding that of the eight estimates, six are above the maximum possible proportion (1.0), further suggests that the upper bound is a generous one; second, that the method is capable of revealing hierarchical structure increases the credibility of the evidence it provides against low-level invariance of multistroke constituents in strings of distinct strokes.

Further Analysis by Segmentation of Single-Stroke Finger Trajectories

The evidence described above argues against the existence of invariant subsequences within strings of distinct strokes, based on the failure of low-level invariance relative to sequence length. Does the invariance property apply in any way to such strings? The small size of the linkage proportion indicates that sequence length influences the *ISI* between essentially every stroke and its successor. Can we find any decomposition of the stream of

---

[20]See Pew and Rosenbaum (1988); Rosenbaum (1985); and Rosenbaum (1987) for such claims, which also cover the production of utterances of unrelated words and of pseudowords. Other features of performance that such a model has difficulty explaining include the observed linearity of the effects of utterance length on latency and mean element duration (Sternberg, Monsell, et al., 1978; Sternberg, Wright, et al., 1980; Sternberg, Knoll, et al., 1989), for which a binary-hierarchy model with constant mean traversal time per node produces decelerating effects, and the small magnitude of the effects on element duration of serial position relative to length.

[21]This observation is consistent with the lower tail of the *ISI* distribution for the doublet being approximately invariant relative to sequence length; we see in a later section that the mean *ISI* is invariant also.

actions between one stroke and the next that reveals invariant segments? The movements associated with each stroke in such a string are complex, and thus susceptible to further analysis. Such analysis requires more information than can be obtained from just the sequence of time points at which keys are actuated, however. We report a study in which we attempted to decompose the trajectories of individual strokes, as we varied the length of the sequence containing them.[22]

The design of one experiment in which we examined strings of consonants typed by the fingers of alternating hands, together with examples of the strings, is shown in Table 1.3. In these examples, the trajectory examined is the one followed by the index finger of the left hand between the strokes corresponding to f and v; between these two strokes the right index finger strikes the m key. Subjects performed such sequences in isolation ($n = 3$), and at all possible serial positions in strings of length $n = 4$ and $n = 5$. The trajectory starts with a *hold* segment, during which the finger depresses a key (f in our example). There follows a *lift* segment during which the finger rises from the key and moves in the plane of the keyboard towards its next target. (It is typically during the lift of one finger that the hold segment of the other fingers' trajectory begins.) Finally there is an *attack* segment, during which the finger continues to move in the plane while it moves down to strike and depress the next key (v in our example). We defined these three segments by means of (low) *velocity* criteria. We used (zero) *acceleration* criteria to further segment the lift and attack into *early* and *late* components, giving us a total of five segments.

Fig. 1.9A shows mean duration of the full trajectory, from the start of the hold through the end of the attack. Because the hands alternate, the trajectory covers two interstroke times. The stroke rate for strings of length $n = 3$, for example, is therefore about 9 strokes/sec. We see the usual effect of sequence length on production rate. Fig. 1.9B shows durations of the five segments. Four of the five segments show the sought-for low-

---

[22]To measure such trajectories we used an apparatus that permits continuous measurement of the tips of two fingers in three dimensions. For each of the fingers to be measured it incorporates a separate lightweight linkage attached to the fingernail and connected to three low-torque potentiometers. The moment-to-moment settings of the potentiometers are uniquely determined by the position of the fingertip in three dimensions. The signal from each potentiometer was filtered, sampled at 200Hz, digitized, and stored for later analysis. We found that attachment of the linkages to two index fingers slowed the mean typing rate by only 4% and induced no change in error rate; the measurement procedure thus appears not to destroy the behavior under investigation. The analysis we present here (Sternberg, Knoll, et al., 1983) was restricted to the vertical component of the trajectories of the two index fingers.

TABLE 1.3.

Examples of letter sequences used for analysis of index-finger trajectories in typing. The examples shown were selected to permit assessment of effects of sequence length on durations of trajectory segments of the left index finger.

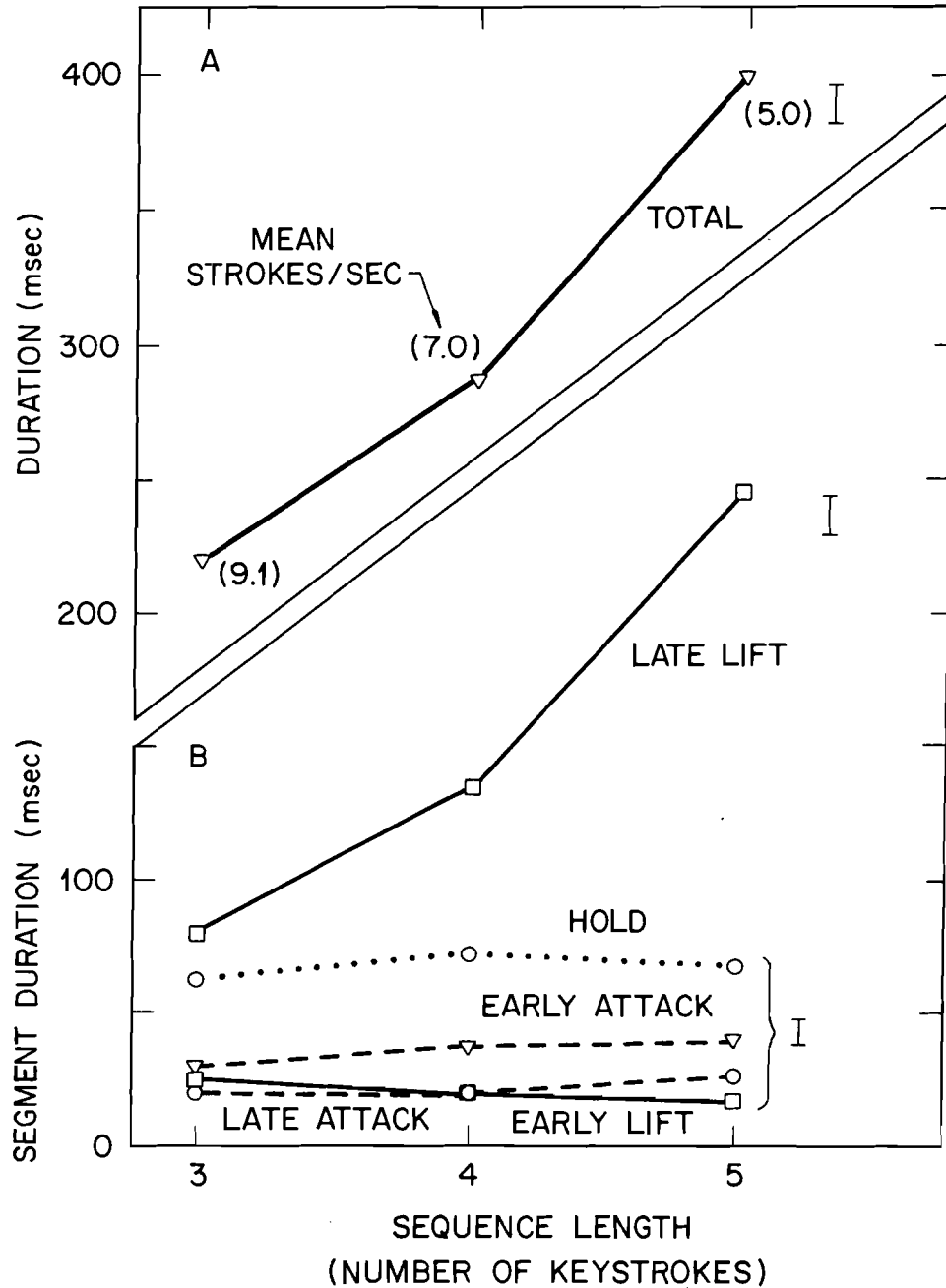| Sequence Length	Other Finger(s)		Measured Index Fingers			Other Finger(s)	
	Left	Right	Left	Right	Left	Right	Left
$n = 3$			f	m	v		
$n = 4$			f	m	v	k	
		p	f	m	v		
$n = 5$			f	m	v	l	s
		k	f	m	v	p	
	d	p	f	m	v		

FIG. 1.9. Trajectory analysis for sequences of keystrokes produced by alternating hands. A: Mean total duration of the time interval between two strokes typed by the same index finger, and estimates of $\pm SE$. [Three subjects; about $18(n - 2)$ observations per point, where n = sequence length.] Because these two strokes are separated by another stroke typed by the other hand, this duration is the sum of two ISIs. B: Mean duration of five trajectory segments whose sum is the total duration shown in Panel A, and estimates of $\pm SE$.
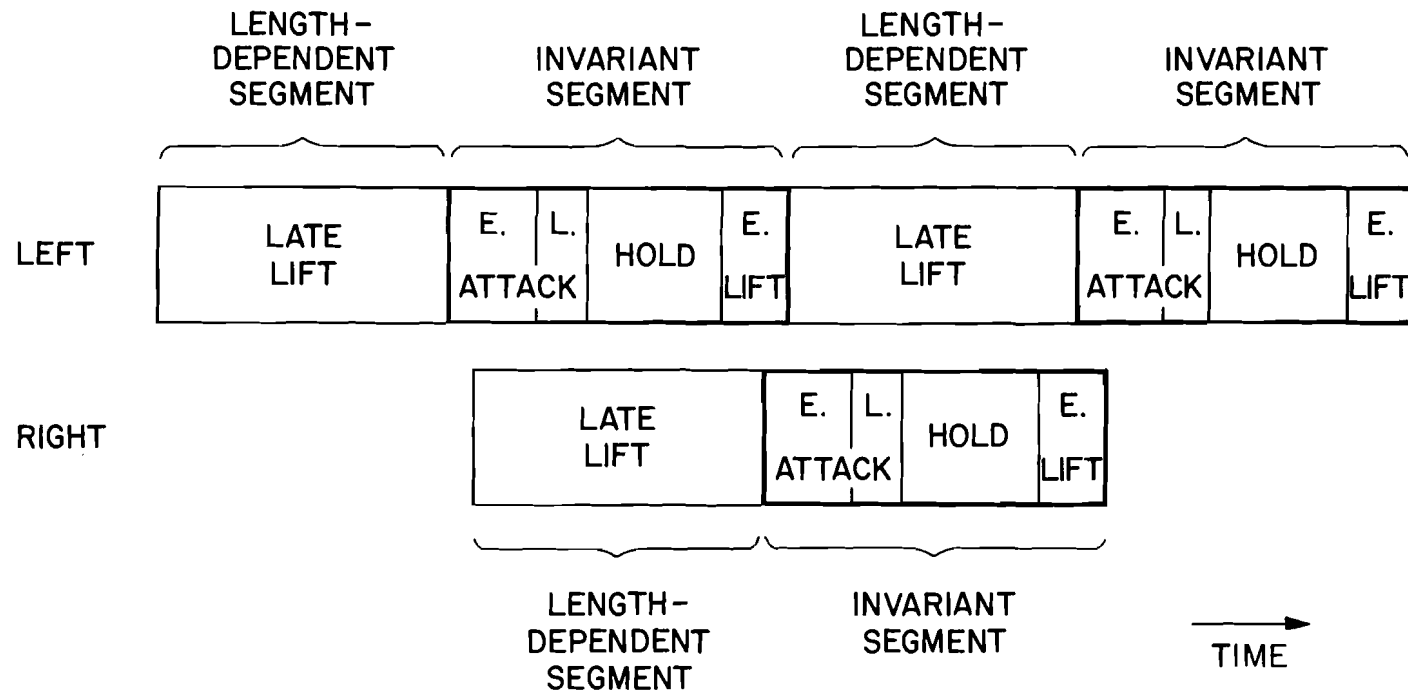
FIG. 1.10. Temporal overlap of finger-trajectory segments in a typed sequence produced by alternating hands. While one hand produces segments whose durations are independent of sequence length (*early attack, late attack, hold,* and *early lift*), the other hand produces its length-dependent segment (*late lift*).

level invariance at the level of individual strokes; only the late lift shows any substantial effect of sequence length.[23]

This promising analysis of segment durations thus reveals a qualified form of low-level invariance at the level of individual strokes, similar to what we have already described for speech; together with the near-zero value of the linkage proportion, it also indicates that a separate action unit is associated with each stroke. It is tempting to argue that as in the case of speech, we need only elaborate a process with transfer of control by adding a mechanism that permits actions controlled by a subprogram to outlast the execution of that subprogram, so as to provide behavioral continuity. The situation in this case is more complicated, however, as indicated in Fig. 1.10. While the index finger of one hand is in its late lift phase (whose duration depends on n, presumably because it occurs while higher level processes select and invoke the next subprogram for that hand), the other hand executes other segments (whose durations are independent of n). With each stroke, then, as the hands alternate, a high-level control process for one hand (sensitive to n) proceeds concurrently with execution of a low-level subprogram for the other (not sensitive to n), producing alternation of autonomously executed subprograms. One possibility is that we may have two separate but concurrent (and coordinated) hierarchical control processes, one per hand, each operating by transfer of control. Other data, however, show that the timing of each hand is influenced by the total number of strokes in a sequence and not just by the number of strokes assigned to that hand, which suggests an integrated control structure for two-hand sequences.[24] If we wish to preserve the idea that this structure is hierarchical, we are forced to conclude that rather than using transfer of control, the associated process functions by command/signal transmission, which, as discussed earlier, permits operations at different levels to proceed concurrently.

Invariance Test of Embedded Doublets

For our final test of low-level invariance we turn to the experiment mentioned earlier on the performance of keystroke strings containing a doublet. As shown by the examples in Table 1.4 we used strings of lengths $n = 2$,

---

[23]In unpublished work we have replicated this finding, and also extended it to strings of letters typed by the fingers of one hand. We have also found that for another influencer—serial position—the same segments are again duration-invariant. That is, the effect of serial position on segment duration is also restricted to the late lift. In this last respect these findings in typing differ from those obtained in speech production, discussed earlier.

[24]These data arise in the production of letter strings that consist of two sucessivley executed substrings, one typed by the fingers of each hand.

TABLE 1.4.
Examples of letter strings containing the embedded doublet *cc,* and control strings.

Sequence Length	Doublet Strings Serial Position of Doublet				Corresponding Control Strings				
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*5*
$n = 2$	cc				cf	xc			
$n = 3$	ccf	xcc			czf	rcz	zsc		
$n = 4$	cczf	rccz	zscc		csrz	zcfx	tscz	szfc	
$n = 5$	ccsrz	zccfx	tsccz	szfcc	cxzfs	tcvsz	zfcvx	fxzcs	szxtc

3, 4, and 5 typed by the fingers of one hand; for each length, each of several doublets (such as *cc*) appeared in each possible serial position; distinct-letter strings (examples in Table 1.4) were among the controls. We reported earlier on our use of the *ISI* distributions for doublets in strings of lengths $n = 2$ and $n = 5$ to test our estimation method, which suggested that the linkage probability for doublets was substantially higher than for nondoublet *ISI*s. Other analyses of the production of strings that include a doublet indicate that the two strokes of the doublet are contained in the same action unit, when the definition of *unit* is provided by a model of rapid sequence production (Sternberg, Knoll, et al., 1983). It follows that the doublet should display low-level invariance. In Fig 1.11 we provide a test, based on the mean *ISI* rather than its distribution: The element duration function for control strings (with a slope of 29 ± 6 msec/letter) is compared to the mean *ISI* for doublets (with a small and nonsignificant slope of 2 ± 3 msec/letter), which we thus find to be essentially invariant relative to string length, the first persuasive instance of low-level invariance above the level of the individual stroke in typing or tapping.

Insofar as the doublet is a low-level constituent, we expect its duration also to be invariant relative to serial position. Consistent with this expectation we find the effects of serial position on doublet duration to be small and nonsignificant at all lengths ($n = 3, 4, 5$).[25]

If the doublet is an action unit, then adding an immediate repeat of one of the strokes in a string of *n* distinct consonants increases the number of strokes to $n + 1$ without increasing the number of *units*. This observation leads to the tests of high-level invariance that we consider later.

Summary of Evidence For and Against Low-Level Invariance

In the preceding sections we have reported on our search for duration-invariance of subsequences of elements and of element segments, relative to sequence length and serial position, where the elements are keystrokes and stress-groups.

In speech production, we find low-level invariance relative to sequence length at the level of the stress group. To make this claim, however, we must modify the invariance property to permit occasional epochs for which invariance fails, a deviation that we explain by the smoothing of action between the end of one unit and the beginning of the next. That serial

---

[25]Significance levels for effects of serial position on nondoublet bigrams ($n = 3, 4, 5$) are (*ns*, .001, .003) in distinct-letter strings, and (.006, .05, *ns*) in doublet strings. Significance levels for the *difference* between effects of serial position on doublet and non-doublet bigrams are (*ns*, .01, .001) and (.02, *ns*, *ns*) for distinct-letter and doublet strings, respectively.
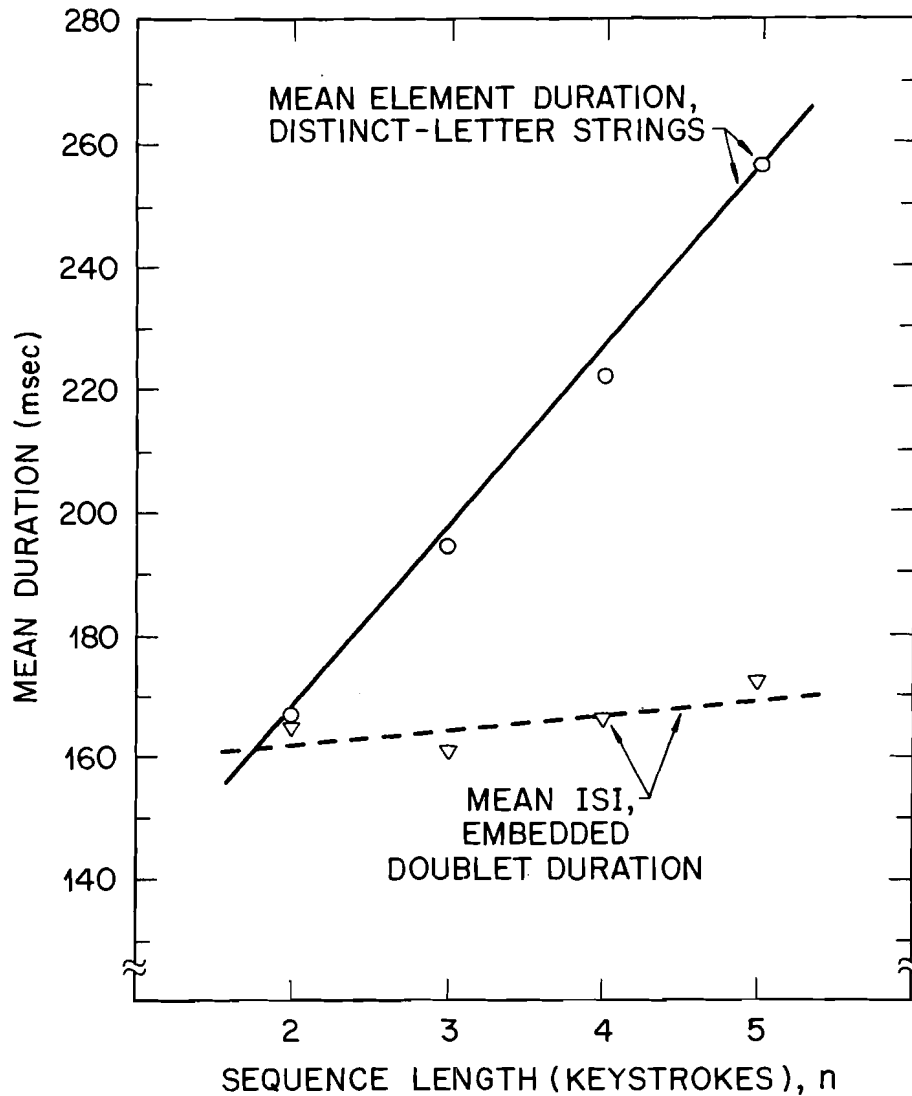
FIG. 1.11.   Mean element duration for typed sequences of distinct letters [fitted function 110 + 29n msec; eight subjects; about 96(n − 1) observations per point; slope estimate SE = 6 msec] and mean doublet duration as a function of the length of the sequence in which it is embedded [fitted function 157 + 2n msec; about 96(n − 1) observations per point; slope estimate SE = 3 msec]. See Footnote 9.

position has an effect on the very segments that are invariant relative to sequence length is a serious challenge, however; this violation could perhaps be dealt with by invoking a process of temporal modulation, possibly related to the imposition of prosodic organization, that operates between a hierarchical control structure and the performance itself. Unless we can specify constraints on such a process, however, the invariance property

loses much of its power. In speech production we have not yet searched for low-level invariance above the level of the stress-group.

In the typing of strings (such as *jfjfkdkd*) that contain repeated stroke-pairs, we find evidence against invariance at the level of the pair relative to both sequence length and serial position, and an indication of failure above that level as well. It would be especially interesting if further work strengthens these findings, given that performance of such sequences has been taken as a paradigmatic case of hierarchical control, based on the pattern of *ISI* means for strings of length $n = 8$. For strings (such as *zccfx*) that contain a repeated stroke (doublet), we find low-level invariance of the doublet itself. Finally, at the level of the single stroke in strings of distinct consonants we find duration invariance of segments of the stroke trajectory, relative to both sequence length and serial position; this supports the invariance property, again modified as mentioned above.

Insofar as low-level duration invariance is regarded as a critical property of action sequences produced under hierarchical control, we thus have evidence for shallow hierarchies: At this writing, only the doublet experiment provides clear evidence for constituents that incorporate more than one stroke (in typing or tapping) or one stress-group (in speech).

TESTS OF HIGH-LEVEL INVARIANCE

Sequence Duration Under Singlet-Doublet Exchange

Both tests of high-level invariance that we describe make use of the doublet experiment, and both assume constraints on the extent of temporal overlap of high-level and low-level operations equivalent to that provided by a process with transfer of control. With respect to the *ISI* of the doublet itself, we provided evidence in earlier sections that neither the lower tail of its distribution nor its mean depends on the number of strokes in the string in which it is embedded, and also that its mean is invariant with respect to serial position. These findings support the hypothesis that the doublet is executed by a subprogram that functions autonomously relative to the control structure that invokes it, as we would expect if that structure were hierarchical.

Combined with high-level invariance, the hypothesis leads to the following complementary property: If we replace a single stroke (*singlet*) in a string of length n by a doublet, creating a string of length $n + 1$, we should have exchanged one autonomous subprogram in the control structure (associated with the singlet) by another one (associated with the doublet), without changing other aspects of the structure. In the simplest case,

the exchange would leave all the *ISI*s unaltered, merely inserting the invariant doublet *ISI* among them. Because the execution of the subprogram may start before the timepoint of its (first) stroke, and may end after the timepoint of its (last) stroke, the exchange might also affect the adjacent *ISI*s. Any such effects, however, should be independent of the embedding structure, and hence of the number of additional strokes $(n - 1)$ and of the serial position of the replacement.

One implication is that the exchange should alter mean string duration by a constant, independent of n: If D_n is the mean duration of a string of n distinct strokes, and D_n^* is the mean duration of a string of n strokes containing an embedded doublet, then we expect that

$$D_n^* = D_{n-1} + k, \qquad (n \geq 2). \tag{6}$$

That is, replacement of singlet by doublet should leave invariant, up to a possible additive constant, the *sum* of the mean *ISI*s for stroke pairs that are common to the two strings (as well as adding the constant mean *ISI* of the doublet).

There are conditions under which validation of Equation 6 would be of little interest, however: Suppose that augmenting a string by one stroke incremented its mean duration by the same constant, regardless of n and of whether the string contained a doublet. Then D_n^* and D_n would be linear with equal slopes, and Equation 6 would follow, as would Equation 7:

$$D_n^* = D_n + c, \qquad (n \geq 2), \tag{7}$$

where c is a different additive constant. Equation (7), however, would be expected to obtain if the doublet functioned simply as a stroke pair with an unusual *ISI*, rather than having the same status in a hierarchical control structure as a singlet. It is by virtue of the nonlinearity of D_n that it is possible to determine that only one of these relationships obtains.

A second implication of doublet-singlet equivalence is that the exchange should leave the *individual* mean *ISI*s invariant, up to possible additive constants, for those stroke pairs adjacent to the replacement. In this section we report a test of the first implication; in the next section, the second.

Results of a test of the first implication are shown in Fig. 1.12. The leftmost function is D_n $(2 \leq n \leq 5)$, the sequence-duration function for non-doublet strings. Translated one unit to the right, it becomes the D_{n-1} function, $(3 \leq n \leq 6)$. The leftmost broken curve is D_n^* $(2 \leq n \leq 5)$, the sequence-duration function for doublet strings. This function is parallel to the D_{n-1} function (i.e., vertically displaced, as described by Equation 6) and not to the D_n function (Equation 7), in accordance with the implication
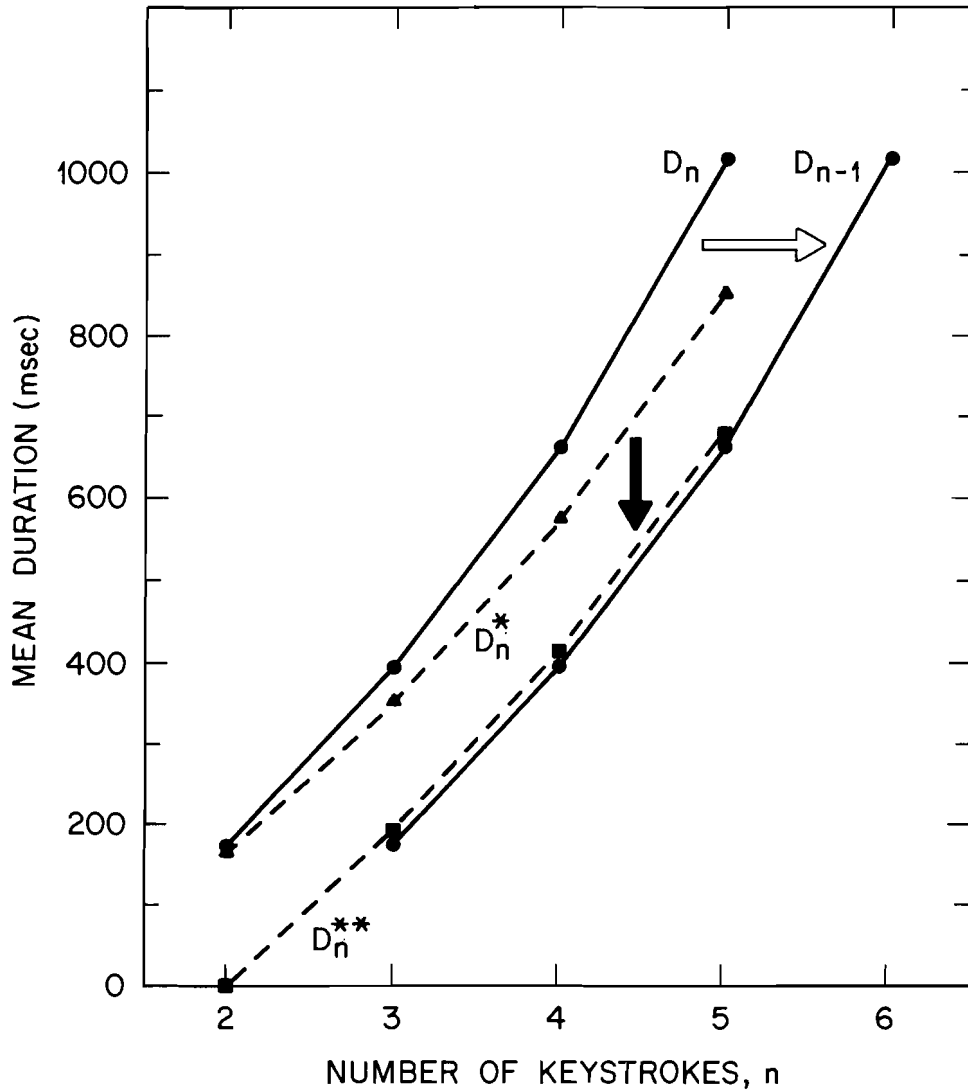
FIG. 1.12.   Sequence durations D_n and D_{n-1} for distinct-letter strings and D_n^* for doublet strings, each containing $2 \le n \le 5$ strokes. Also shown is $D_n^{**} = D_n^* - ISI_n^*$, the sum of the non-doublet ISIs in doublet strings. D_{n-1} is the same as D_n, but translated one unit to the right. See text for explanation.

under test. To obtain the lower of the two broken curves, we have subtracted the mean doublet duration ISI_n^* (shown in Fig. 1.11) for strings of length n to give the mean sum of ISIs for non-doublet stroke pairs in doublet strings, $D_n^{**} = D_n^* - ISI_n^*$. Agreement of D_n^{**} with D_{n-1} is not perfect; apparently there is an effect of the singlet-doublet exchange on contextual ISIs. However, this effect is both small (about 20 msec) and

independent of n.[26] Thus, not only does its context not influence the doublet (low-level invariance), but also, except for a small constant effect, exchanging singlet for doublet does not appear to alter its context (high-level invariance).

Consider the implications of this finding for the kind of control process that underlies the production of one-handed letter strings. Replacing a singlet by a doublet increases the time required for subprogram execution, providing additional opportunity for overlap of high- and low-level processes insofar as such overlap is possible, as in a process with command/ signal transmission. As we see essentially no indication of such overlap, a process with transfer of control (in which no overlap is permitted) easily explains these results.

Serial-Position Effect Under Singlet-Doublet Exchange

For this second test of high-level invariance we performed two different analyses. We exemplify them for a doublet string S_n^* with $n = 5$ strokes, whose doublet is associated with the third ISI: $S_5^* = zxttc$. In the first analysis (comparison by *subprogram position*), the doublet string is compared to a non-doublet string with one less stroke, but with the same number of hypothetical subprograms: $S_4 = zxtc$; we compare the three pairs of $ISIs$ determined by the bigrams zx, xt, and tc that are shared by the two strings. $ISIs$ in corresponding positions should match, if the control structure satisfies high-level invariance as doublet is exchanged for singlet. In the second analysis (comparison by *stroke position*), the doublet string is compared to a non-doublet string with the same number of strokes: $S_5 = zxtfc$; we compare the two pairs of $ISIs$ determined by bigrams zx and xt that are shared by the two strings, and the pair associated with tc in S_5^* and fc in S_5. In contrast to the analysis by subprogram position, invariance provides no reason to expect $ISIs$ in corresponding stroke positions to match, even if we adjust for the difference between overall mean interstroke times.

To perform the analyses we combined data over doublet strings with

---

[26]Analysis of variance shows a significant ($p = .01$) interaction of sequence length n with D_n^* versus D_n, ($2 \leq n \leq 5$), but none ($p > .25$) for D_n^{**} versus D_{n-1}; the latter analysis reveals a significant ($p < .05$) main effect, however. Another approach to the data dramatizes the usefulness of the nonlinerarity of D_n: We find $(D_4^* - D_2^*) - (D_4 - D_2) = -78.6 \pm 13.0$ msec ($p < .001$), where the 7 df SE is based on variability over the eight subjects, whereas $(D_5^{**} - D_3^{**}) - (D_4 - D_2) = 490.7 - 488.4 = 2.3 \pm 6.9$ msec, supporting the remarkable parallelism of means displayed in the figure.

the doublet in different serial positions.[27] Results are shown in Fig. 1.13.

For the stroke-position analysis (upper panel) we had strings of three lengths to consider. In each case (n = 3, 4, 5), the shapes of the serial-position functions being compared were reliably different ($p \approx .001$). When we move from the stroke-position to subprogram-position analysis shown in the lower panel, we find a reliable ($p < .001$) improvement in agreement, in the direction of what is implied by high-level invariance. For n = 5, however, the shapes for doublet and distinct-letter strings are still significantly different ($p \approx .007$), indicating a failure of high-level invariance in the distribution of duration over serial positions.

The exchange of doublet for singlet thus appears to leave the summed durations of other constituents invariant, in accordance with high-level invariance, but alters the way in which the sum is distributed.

JOINT TESTS OF HIGH-LEVEL AND LOW-LEVEL INVARIANCE

Given that the action unit in the rapid articulation of short utterances is the stress group, an experiment in speech production suggests itself that is analogous to the doublet experiment examined above. One could compare utterances of various lengths composed exclusively of monosyllabic units with the same utterances after replacing one of the units by a multisyllabic unit. To our knowledge, no such experiment under rapid-speech conditions has been performed. However, there are studies of both typing and speech production in which each string is homogeneous with respect to unit size (in syllables/unit in speech, for example) and in which pairs of strings can be regarded as being related by exchange of all units for others of a different size. (For example, consider bay-rum-limb and baby-rumble-limit.) For such experiments, the combination of high-level and low-level invariance properties, given a process with transfer of control, has a surprisingly strong implication, which can thus be used to test these properties in combination.

---

[27]The combination rule differed for the two analyses. For the analysis by subprogram position, we assigned positions to ISIs by ignoring the doublet ISI; for the doublet string S_5^- = zxttc the nondoublet ISIs would then be numbered 1, 2, and 3. We then averaged correspondingly numbered ISIs within lengths. For the analysis by stroke position, we assigned positions to ISIs that took into account the presence of the doublet ISI; for S_5^- = zxttc the nondoublet ISIs would then be numbered 1, 2, and 4. We excluded the doublet ISI itself from the analysis, and again averaged correspondingly numbered ISIs.
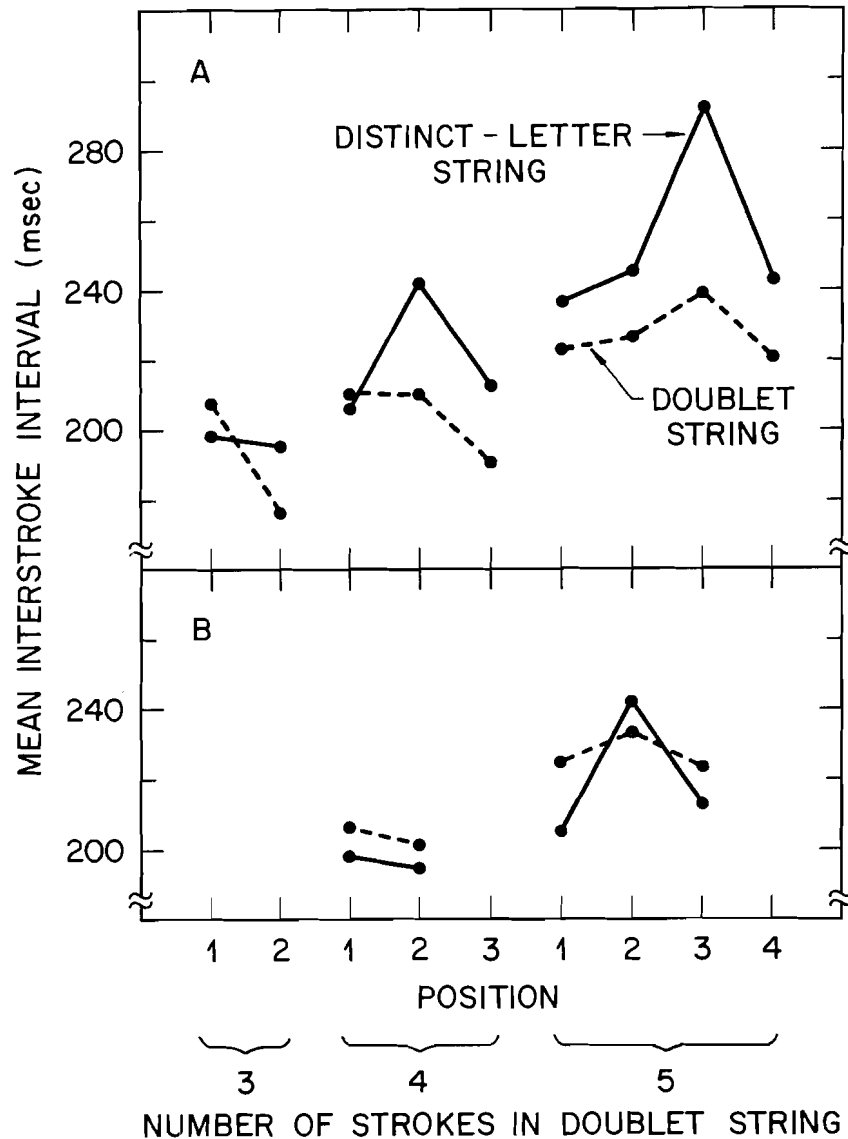
FIG. 1.13.   Serial-position effects in distinct-letter strings and doublet strings of three lengths. A: Mean *ISIs* matched by stroke position. B: Mean *ISIs* matched by subprogram position.

We develop the argument with respect to sequence duration. Given high-level invariance, we can replace one unit (u_x) of an n-unit string by another (u_y), leaving the remainder of the control structure invariant. Because the execution times of u_x and u_y differ, in general, such an exchange would alter sequence duration. Low-level invariance implies that the execution times, and hence their difference, are independent of the higher level control structure, and hence are independent of sequence

length, n. We can therefore describe the difference as a constant rather than a function of n: Let u_y be the unit with the longer execution time, and let δ_{xy} represent the mean execution-time difference. Given a process with transfer of control, δ_{xy} would also describe the mean sequence-duration difference due to the exchange. High-level invariance implies that we can make additional exchanges, until all n u_x-units have been replaced by u_y-units, again leaving the remainder of the control structure invariant, and thus generating a total sequence-duration increment $n\delta_{xy}$. The difference between the durations of sequences of n u_x-units and n u_y-units can thus be at most linear in n:

$$D_n(u_y) \; - \; D_n(u_x) \; = \; n\delta_{xy}. \tag{8}$$

It follows that *any nonlinear term in the sequence-duration functions $D_n(u_x)$ and $D_n(u_y)$ must be the same, regardless of the sizes of units u_x and u_y.* If the duration functions are quadratic, as they appear to be, their quadratic coefficients must therefore be invariant under unit exchange. Equivalently, the mean element-duration functions, $d._n(u_x)$ and $d._n(u_y)$, must have equal slopes.

Effects of Utterance Length Under Speech-Unit Exchange

Results of one experiment, with unit size 1 versus 2 versus 3 syllables, and typical results, were described in an earlier section. There we reported that a plausible unit definition (the stress group) led to mean element-duration functions with equal slopes (Fig. 1.1, Table 1.1). Similar results have been reported for an experiment comparing utterances of monosyllabic and disyllabic nouns (Sternberg, Monsell, et al., 1978). Because such evidence favors an implication of the two invariance properties jointly, together with a process of transfer of control, it lends support to the properties and process.[28]

Effects of Number of Strokes Under Keystroke Exchange

For skilled transcription typing it is well known that *ISI*s are shorter between strokes typed by different hands than by the same hand (Gentner, 1983). For a letter string transcribed by one hand the duration is thus greater than for a string transcribed by alternating hands; this is also true for strings prescribed and prepared in advance. If we assume that the action

---

[28]It also follows from the present argument that the shapes of serial-position functions for utterances with a specified number of units should be independent of unit size. This implication has not yet been tested in speech production.

unit for both hand patterns is the single stroke, then the difference in interstroke transition perhaps corresponds to a difference in unit size, and the relation between the two element-duration functions is of special interest. The functions in one comparison were essentially parallel, which again supports the invariance properties; findings on this point are mixed, however.[29]

Summary of Evidence For and Against High-Level Invariance

Fewer data are available that permit testing high-level than low-level invariance, but here, also, results are mixed. In both typing and speech production, where we have compared sequences in which constituents of different size are exchanged, we find striking support for high-level invariance (together with transfer of control) in measures of sequence duration. But where we have examined the distribution of sequence duration over serial positions (in typing) we find that it is altered by the exchange. Again, as in our suggestion of how low-level invariance might be salvaged given disallowed effects of serial position, advocates of hierarchical control may be forced to argue for a process of temporal modulation that intervenes between the output of a hierarchical control structure and the actions it controls.

CONCLUSIONS

High-level and low-level invariance seem to us to be inherent in the idea of hierarchical control in the execution of action sequences. Furthermore, we have found instances in both speech and typing where these properties apply, especially if we allow for the smoothing of action at constituent

---

[29]Sternberg, Monsell, et al. (1978) reported element-duration slopes of 15.2 and 14.1 msec/stroke for alternating- and one-hand strings, respectively. Although we have replicated this finding, we have also observed violations, with a tendency for a larger slope in the production of alternating-hand strings; more work is needed. Note also that even in the study mentioned, the magnitude of the effect of serial position is larger for alternating-hand strings, instead of being invariant as required. A second difficulty arises from the analysis of finger trajectories as described, which suggests that the change from one-hand to alternating-hand strings is more complicated than merely a change of unit size, because it requires concurrent processes at different levels of control. Indeed, this analysis induces skepticism about the finding of parallel element-duration functions, and suggests attempting alternative manipulations of unit size in typing.

boundaries. Within the limited domain of duration invariance there is also a pattern of violations, however, associated with the distribution of duration over serial position—violations that suggest that significant control operations must intervene between the output of a hierarchical structure and the actions it controls.

In our limited search we have found little evidence for anything other than shallow hierarchies, even though our procedures were designed to maximize both the opportunity and the incentive for advance planning. Until further work is done, including the use of other measures and the study of other factors, it seems fair to conclude that the largest constituents smaller than the whole sequence in rapid utterances or keystroke strings are seldom larger than the elements we have described as action units—individual strokes in typing and stress groups in speech.

In much of our analysis we have assumed that the process that operates on the hierarchical structure does so by transfer of control, and some of our findings in both speech and one-hand typing strongly support this assumption. Our measurements also suggest, however, that a different control process must be operating when coordination of the hands is required, one that permits concurrent operations at different levels, such as control/signal transmission.

Among the questions our findings suggest are the following: Are there conditions of sequence production not yet considered that would reveal deeper hierarchies? What other interesting properties might a hierarchical control structure confer on the sequences it produces? Are there other conceptions of hierarchical control in the execution of action sequences to which the invariance properties do not apply, but that are nonetheless testable?

ACKNOWLEDGMENTS

REFERENCES

Abend, W., Bizzi, E., & Morasso, P. (1982). Human arm trajectory formation. *Brain, 105*, 331–348.

Aho, A. V., Hopcroft, J. E., & Ullman, J. D. (1983). *Data structures and algorithms.* Reading, MA: Addison Wesley.

Barlow, G. W. (1977). Modal action patterns. In T. A. Sebeok (Ed.), *How animals communicate* (pp. 98–134). Bloomington: Indiana University Press.

Benecke, R., Rothwell, J. C., Day, B.L., Dick, J. P. R., & Marsden, C. D. (1986). Motor strategies involved in the performance of sequential movements. *Experimental Brain Research, 63*, 585–595.

Book, W. F. (1908). *The psychology of skill with special reference to its acquisition in typewriting.* Missoula, MT: University of Montana.

Broadbent, D. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181–201.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*, 345–375.

Collard, R., & Povel, D.-J. (1982). Theory of serial pattern production: Tree traversals. *Psychological Review, 89*, 693–707.

Dawkins, R. (1976). Hierarchical organization: A candidate principle for ethology. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 7–51). Cambridge: Cambridge University Press.

Dell, G. S. (1984). The representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 222–233.

Drummond, H. (1981). The nature of and description of behavior patterns. In P. P. G. Bateson & P. H. Klopfer (Eds.), *Perspectives in ethology: Vol. 4.* New York: Plenum.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Fendrick, P. (1937). Hierarchical skills in typewriting. *Journal of Educational Psychology, 28*, 609–620.

Fentress, J. C. (1983). Ethological models of hierarchy and patterning of species-specific behavior. In E. Satinoff & P. Teitelbaum (Eds.) *Handbook of behavioral neurobiology* *(Vol. 6)* (pp. 185–234). New York: Plenum.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica, 38*, 35–50.

Fowler, C. A. (1985). Current perspectives on language and speech production: A critical review. In R. Daniloff (Ed.), *Speech science: Recent advances* (pp. 193–278). San Diego: College-Hill.

Fromkin, V. A. (Ed.). (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.

Fromkin, V. A. (Ed.). (1981). *Errors of linguistic performance: Slips of the tongue, ear, pen, and hands.* New York: Academic.

Gallistel, C. R. (1980). *The organization of action: A new synthesis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. R. (1983). Keystroke timing in transcription typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typing* (pp. 95–120). New York: Springer-Verlag.

Gentner, D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review, 94*, 255–276.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15,* 411–458.

Gordon, P. C., & Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *Journal of Memory and Language, 26,* 300–321.

Greeno, J. G., & Simon, H. A. (1974). Processes for sequence production. *Psychological Review, 81,* 187–198.

Jagacinski, R. J., Marshburn, E., Klapp, S. T., & Jones, M. R. (1988). Tests of parallel vs. integrated structure in polyrhythmic tapping. *Journal of Motor Behavior, 20,* 416–442.

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. In G. H. Bower (Ed.) *The psychology of learning and motivation (Vol. 4)* (pp. 172–247). New York: Academic.

Keele, S. W. (1987). Sequencing and timing in skilled perception and action: An overview. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 463–487). London: Academic.

Klein, R. (1983). Nonhierarchical control of rapid movement sequences: A comment on Rosenbaum, Kenny, and Derr. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 834–836.

Lehrdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception, 1,* 229–252.

Leonard, J. A., & Newman, R. C. (1964). Formation of higher habits. *Nature, 203,* 550–551.

Machlis, L. (1977). An analysis of the temporal pattern of pecking in chicks. *Behaviour, 63,* 1–70.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–506.

Massey, J. T., Schwartz, A. B., & Georgopoulos, A. P. (1986). On information processing and performing a movement sequence. *Experimental Brain Research (Supplement) 15,* 242–251.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehardt, & Winston.

Monsell, S., & Sternberg, S. (1981). *Speech programming: A critical review, a new experimental approach, and a model of the timing of rapid utterances (Part 1).* Bell Laboratories Technical Memorandum, June. Murray Hill, NJ: AT&T Bell Laboratories.

Morasso, P. (1983). Three dimensional arm trajectories. *Biological Cybernetics, 48,* 187–194.

Morasso, P., Mussa Ivaldi, F., & Ruggiero, C. (1983). How a discontinuous mechanism can produce continuous patterns in trajectory formation and handwriting. *Acta Psychologica, 54,* 83–98.

Narmour, E. (1983). Some major theoretical problems concerning the concept of hierarchy in the analysis of tonal music. *Music Perception, 1,* 129–199.

Nelson, K. (1973). Does the holistic study of behavior have a future? In P. G. Bateson & P. H. Klopfer, *Perspectives in Ethology* (pp. 281–328). New York: Plenum.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Pew, R. W., & Rosenbaum, D. A. (1988). Human movement control: Computation, representation, and implementation. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.) *Stevens' handbook of experimental psychology, 2nd edition: Vol. 2. Learning and cognition* (pp. 473–509). New York: Wiley.

Perkell, J. S. (1980). Phonetic features and the physiology of speech production. In B. Butterworth (Ed.), *Language production: Vol. 1, Speech and talk* (pp. 135–173). London: Academic.

Reitman, J. S., & Reuter, H. H. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology, 12,* 554–581.

Rosenbaum, D. A. (1983). Hierarchical versus nonhierarchical models of movement sequence control: A reply to Klein. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 837–839.

Rosenbaum, D. A. (1985). Motor programming: A review and scheduling theory. In H. Heuer, U. Kleinbeck, & K.-H. Schmidt (Eds.), *Motor behavior: Programming, control, and acquisition* (pp. 1–33). Berlin: Springer-Verlag.

Rosenbaum, D. A. (1987). Hierarchical organization of motor programs. In S. P. Wise (Ed.), *Higher brain functions: Recent explorations of the brain's emergent properties* (pp. 45–66). New York: Wiley.

Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 86–102.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6,* 1–36.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109–136). New York: Springer-Verlag.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106,* 467–482.

Sternberg, S., & Knoll, R. L. (1986). *Determination of the unit in rapid action sequences.* AT&T Bell Laboratories Technical Memorandum. Murray Hill, NJ: AT&T Bell Laboratories.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1983). *Control of rapid action sequences in speech and typing.* Bell Laboratories Technical Memorandum. Murray Hill, NJ: AT&T Bell Laboratories.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1989). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45,* 175–197.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *The perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Lawrence Erlbaum Associates.

Szentagothai, J., & Arbib, M. A. (1975). *Conceptual models of neural organization.* Cambridge, MA: MIT Press.

Terzuolo, C. A., & Viviani, P. (1980). Determinants and characteristics of motor patterns used for typing. *Neuroscience, 5,* 1085–1103.

Volpert, W. (1982). The model of the hierarchical-sequential organization of action. In W. Hacker, W. Volpert, & M. von Cranach (Eds.), *Cognitive and motivational aspects of action* (pp. 35–51). Amsterdam: North Holland.

Vorberg, D., & Hambuch, R. (1978). On the temporal control of rhythmic performance. In J. Requin (Ed.), *Attention and Performance VII.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Vorberg, D., & Hambuch, R. (1984). Timing of two-handed rhythmic performance. In J. Gibbon & L. Allan (Eds.), *Timing and time perception. Annals of the New York Academy of Sciences, 423,* 390–406.

Wall, R. (1972). *Introduction to mathematical linguistics.* Englewood Cliffs, NJ: Prentice-Hall.

West, L. J. (1969). *Acquisition of typewriting skills*. Belmont, CA: Pitman.

Wilson, D. (1969). Forms of hierarchy: A selected bibliography. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 287–314). New York: Elsevier.

Wing, A. M. (1977). Effects of type of movement on the temporal precision of response sequences. *British Journal of Mathematical and Statistical Psychology, 30,* 60–72.