Visual Cognition and Action

An Invitation to Cognitive Science

Volume 2

---

edited by Daniel N. Osherson,
Stephen Michael Kosslyn, and John M. Hollerbach

The MIT Press
Cambridge, Massachusetts
London, England

List of Contributors

Irving Biederman
Department of Psychology
University of Minnesota

E. Bizzi
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Fred Dretske
Department of Philosophy
Stanford University

Henrietta L. Galiana
Department of Biomedical Engineering
McGill University

Alvin Goldman
Department of Philosophy
University of Arizona

John M. Hollerbach
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Stephen Michael Kosslyn
Department of Psychology
Harvard University

F. A. Mussa-Ivaldi
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Daniel N. Osherson
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Elizabeth S. Spelke
Department of Psychology
Cornell University

S. Ullman
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

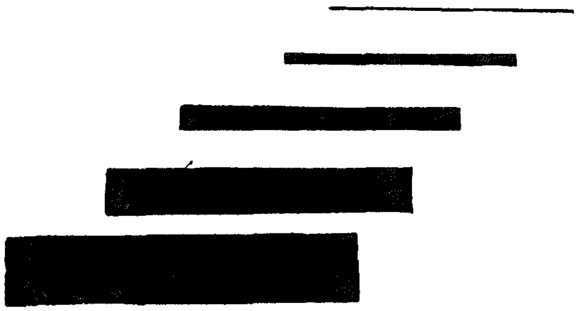
Charles E. Wright
Department of Psychology
Columbia University

A. L. Yuille
Division of Applied Sciences
Harvard University

Chapter 5

Controlling Sequential Motor Activity

Charles E. Wright



5.1 Plans and Planning

To start thinking about the issues covered in this chapter, consider planning and then carrying out a sequence of actions. This might be a high-level plan toward a long-term goal, such as to ensure that you do well in school this semester, or it might be a much more limited and concrete plan, such as organizing a day involving several errands. In the second example, before setting out, you might decide on a tentative plan describing the order in which you will do the errands and how you will get from place to place. This plan is probably largely determined by the location of each errand and the layout of available transportation. As part of this planning process you

I thank my colleagues Ron Knoll at AT&T Bell Laboratories, Stephen Monsell at the University of Cambridge in England, and Saul Sternberg at the University of Pennsylvania for their permission to describe both work that we carried out jointly and, in some cases, work that they have done in which I did not participate. Nina Macdonald and Saul Sternberg provided useful comments on an earlier draft of this chapter. Writing of this chapter was partially supported by the National Science Foundation under grant BNS-87-11273.

will have explicitly made decisions about many details of your plan—what routes to take, where to stop for a break, and so on—while leaving others undecided. You may also have included some contingencies in your plan: for instance, what to do if the bookstore really doesn't open until ten o'clock.

Once the plan exists, you will need some way of recording it for future reference, as well as a mechanism to ensure that its steps are carried out. In a simple case you might just jot the plan on a piece of paper with the idea of referring to it at regular intervals. In a more complex case, such as planning for the semester, you might want to set up a timetable and goals, perhaps using a calendar to ensure that the plan stays on course.

Earlier chapters have examined the planning process for single, isolated movements in various domains. Here we will consider whether, when performing a task that involves sequences of these movements, we plan the sequence in anything like the way we plan a morning's shopping expedition. People often find this suggestion controversial, perhaps because this planning process, if it exists, is one we are rarely aware of. If movement sequences are not planned, however, it is hard to understand how and where we accumulate the information necessary to make them skillfully.

The mechanisms used by the brain to make, store, and carry out plans for sequences of movements have been an open issue for many years. Neurophysiologists, for instance, have only in the last few years begun to understand how, short of growing new physical connections, the nervous system can establish permanent (and/or temporary) links between preexisting pieces of information in the brain. If we are to combine preexisting specifications for unrelated, simple movements to produce more complex concatenations—for example, coordinated head and eye movements to track an object—then some mechanism for producing or simulating such connections or links must be necessary. Further, if the order of the component movements in a movement sequence is critical—a child learning to grasp after reaching, where both individual behaviors are well established—then these links must contain additional information specifying their coordination. This is the problem of *serial order* made famous by Lashley (1951). Finally, in some cases—for instance, playing a musical instrument—the precise timing of successive submovements may be important over and above the requirement that they be produced in a particular order. In this case a movement plan would need to be even more specific.

On the other hand, there are theorists who suggest that these "mental" links between successive actions do not exist. Classical behaviorists, for example, reject the explanation that internal, mental links are established that specify or control the successive components of a movement sequence. Instead, they theorize that sensory stimulation generated by the production of one submovement elicits, by strength of habit, the next

appropriate movement as a response. In this conception, a plan for a movement sequence, rather than being a centralized, cognitive entity with an independent existence, is instead a result emerging from a distributed set of stimulus-response habits. According to this view, the notion that we engage in planning processes that result in independently existing mental plans is simply a misconception based on faulty introspective evidence.

One problem with the behaviorist perspective is that under such an approach, sequencing depends on a feedback process with moderately large delays. These delays establish the granularity of precise movement-sequence timing. In addition, the introduction of feedback delays can exacerbate control problems (see section 1.5.2); even if the stimulus for a response is the efference copy of the previous response—that is, a copy of the outgoing motor commands routed back as a stimulus—the loop delays would probably be on the order of 100+ milliseconds. These problems, along with widespread disenchantment concerning the explanatory power of the behaviorist viewpoint, crystallized for many by the criticisms in Lashley's (1951) article, were a major factor that generated interest in questions about the creation and representation of mental plans. This interest was one of the central themes in the early years of cognitive psychology; one influential monograph of that period was in fact entitled *Plans and the Structure of Behavior* (Miller, Galanter, and Pribram 1960). The nature of planning and plans has also been an important topic in many of the other disciplines that feed into cognitive science: computer science, linguistics, artificial intelligence, and robotics. In this chapter we will explore the nature of plans used to control sequences of simple movements as well as the cognitive mechanism involved in carrying out those plans.

5.1.1 Planning Movement Sequences: Motor Programs

Within the domain of motor control, plans are typically referred to as *motor programs.* Such programs, which take their name by explicit analogy to computer programs, are thought to be involved in activities as diverse as touching an object, hitting a ball with a bat, walking, running, pole-vaulting, driving a car, writing with a pen, and producing speech. They are often thought of as the mental representation that bridges the processes of planning and control (as discussed in section 1.1). More specifically, motor programs are seen as the repository for the accumulated information that underlies skilled, fluent activity. Clearly, a better understanding of the properties of motor programs for movement sequences as well as the planning and control processes that create and interpret them could have far-reaching implications for learning and ultimate performance in many areas.

This topic is also interesting because, although it shares many characteristics with other investigations of cognitive plans and planning—planning a shopping trip, a strategy to solve a physics problem, or a way to put an idea into words—it also differs from them in many important ways. Among other differences, executing a motor program results in overt, physical activity (movements) that can be measured precisely. These movements are usually relatively simple and repeatable. But, at the same time, unlike the way we view some other cognitive plans, we generally do not have the sense that motor programs can be the objects of introspection. The activities involved seem too automatic. For example, I can think about how to throw up a tennis ball and swing a racquet to produce a well-placed serve with top spin. I can read books on the subject or observe others who are experts. As part of this study process, I could learn exactly which muscles are activated, in which sequence, and at what level, as well as the ideal trajectory of the racquet, limbs, and so on. But simply learning this information will not allow me to duplicate the required movement. Thus, although I might somewhat improve my serve with all of this information, it is doubtful whether, without extensive, repetitive practice, I could ever master this movement sequence. In addition, although repetitive practice appears to be necessary for mastery of this skill, it is usually the case that we are unaware of what is changing or what things we are learning as we practice. I have more confidence, however, that I could improve my facility to write and speak more eloquently or to solve physics problems using just such processes of observation and reflection.

5.1.2 Representation of Motor Programs

The information contained in motor programs could potentially be specified at one (or, simultaneously, more than one) of several levels—high-level intentions, endpoint trajectories, joint angles, task dynamics, muscle force distributions, and so on—with necessary missing information presumably computed during the course of a movement. In particular, a number of theorists have proposed that motor programs are learned and organized hierarchically, much as subprograms of a computer program might be organized, and written, in a hierarchical fashion (Greene 1972; Rosenbaum 1985; Saltzman 1979).

Consider, for example, how an infant learns to reach through space and to grasp an object when it touches her hand. What might this learning consist of? How are these newly acquired movements represented so they can be repeated? Later, these activities are refined and combined so that reaching for and grabbing an object becomes a single fluid motion. Later still, this combination may be integrated with locomotion to grab objects that are out of reach. Finally, these same activities may be further refined and combined with other separately learned sequences to allow an infielder in baseball to instinctively (since, presumably, there is no time to create a new plan) and fluidly dive for a hard-hit ground ball, catch it, roll, and come up throwing the ball to first base. In what sense are these complex, learned activities made up of previously learned components? How does being integrated into a higher-level activity change the components?

In this chapter we will not specifically explore these questions of level representation and the hierarchical nature of motor programs. However, it is important to remain aware of these issues, since they will never be far below the surface.

5.1.3 Confirming the Existence of Motor Programs

Before attempting to study the structure of motor programs or the mechanisms used to sequence their elements, we need criteria that determine when the movements in a sequence are made according to a plan. The criteria we use must distinguish the intention to make a movement sequence, couched in terms of abstract, high-level goals, from a plan that is sufficiently detailed to control the neuromuscular system during the performance of that sequence—eventually directing muscles to contract and effectors to rotate about their joints. For example, a plan for typing a sequence of keys on a keyboard (one task that we will discuss extensively later) would presumably include motoric details about which fingers to use, which way they should move, and in what order their movements should be initiated, along with, or instead of, information about which letters are to be typed.

In addition to expecting a motor program to contain "motoric" details, we might also expect this representation to exist prior to the start of the movement.[1] In particular, given this chapter's emphasis on movement sequences, our criteria for performances controlled by motor programs should allow us to discriminate between these and performances produced without planning. One example of the latter would be response chaining, such as posited by behaviorists, where feedback from one action initiates the next action. In this scheme, the transition from one component action to the next depends only on local factors—that is, the particular previous action and the subsequent action that feedback from the first action elicits—not on any knowledge about other actions occurring later (or earlier) in the sequence.

Consider, for example, the sequence of movements required to draw a figure. If I draw this figure by making a stroke, stopping to compare the result with what I wish the overall result to be, choosing another stroke to

---

[1] This expectation that "plans" exist prior to the beginning of a movement should not be taken to exclude the possibility of limited computations taking place during a movement that might alter the outcome.

follow the first, making this second stroke, and so on, an observer might be inclined to reject this as an example of a performance in which the sequence of movements (strokes) is governed by a motor program. And yet in this case there is clearly at least some high-level goal: I have an idea of the figure I wish to draw. To get a feel for the difference between these two processes, try writing cursively the words *motor program* first with your dominant hand (your right hand, if you are right-handed) and then with your nondominant hand. For most people, dominant-hand writing is fluent, quick, and almost effortless, whereas nondominant-hand writing takes much more time, is jerky, and requires substantial attention and concurrent visual feedback. And yet, the products of these two subjectively dissimilar processes are often quite similar and recognizably due to one person. This difference usually becomes more obvious if the movements are made without concurrent visual feedback, as would be the case if your eyes were closed.

From this distinction, we should not draw the conclusion that motor programs always proceed "open-loop," without making use of feedback. Instead, the role of feedback in performing a programmed movement sequence probably is to fine-tune the ongoing performance rather than to aid in the selection (or chaining) of subsequent submovements in the movement sequence. Another potential function of feedback is to allow the motor-system controller to gather information that can be used to update its model of the muscles and the load, allowing compensation for effects of muscle fatigue, poorly estimated load characteristics, and deficiencies in motor programs. Better information about any of these aspects of a movement should improve the performance during subsequent movement sequences.

Devising criteria to distinguish the various possible representations and processes that may underlie the production of a movement sequence is made more difficult by the layered nature of the motor system (see chapter 1). At the lowest level are the immediate, albeit passive, responses to changes in load that result from the stiffness of muscle itself or the nonlinear relations between muscle force, neural activation, muscle length, and contraction velocity. Moving up through the neural control system, there is active feedback from spinal reflexes, pattern generators, and transcortical feedback pathways. Because of these and other mechanisms, there are, interposed between a plan and the resulting movement, myriad potential sources of movement modification and control. Though it may be, as some argue, that the actions of these peripheral layers of the motor system make high-level control feasible, from the perspective of studying the higher-level processes, these peripheral processes are a great source of complexity and confusion. In effect, these peripheral processes interpose unknown layers of buffering and filtering between us, as observers, and the central processes that plan or control movement sequences. Of course, this situation is hardly unique to the study of motor control or cognitive psychology in general. Unfortunately, there are no widely accepted rules that tell us what data will support particular inferences about motor programs.

In the search for movement-sequence regularities, some regularities will be more helpful than others in establishing the existence and nature of the movement-planning process. We should be especially interested in details of the performance at an early point in a movement sequence that depend on the nature of some later part of the sequence. Such observations are of particular interest since they fit well with our intuitions of a planned sequence of movements and would be difficult to account for in a sequence of movements that proceed without a plan. The longer the span of this influence, the more strongly it suggests that the influence results from a planning process rather than a local interaction of sequence elements.

The problem of local interactions would not be a concern if we could assert that one movement in a sequence does not begin until the previous movement has ended. It is possible that successive submovements in a sequence can overlap to some degree—that is, some commands related to a subsequent element in a sequence may be issued while an earlier sequence element still has primary control of the movement effector(s). These anticipatory (or perseveratory) commands will change the path of the movement effector(s) during the period of overlap. Looking from the outside in, it is difficult to distinguish whether these changes are the result of a mechanism, perhaps in the peripheral motor system, acting locally (in the terminology of chapter 1, an effect of the control process) or the result of a central planning process.

This overlap phenomenon and the interpretational difficulties associated with it have been particularly important in theorizing about the production of speech, where such overlap is often referred to as *coarticulation* (Kent and Minifie 1977). The speech sound stream is often represented as a linear sequence of abstract sound segments chosen from the small, discrete set of phonemes of a language. Attempts to locate the boundaries of these segments in articulatory events are usually confounded, however, by the fact that the articulatory movements appear to overlap one another in a complex fashion. This overlap also appears to make the details of the sound structure of a phoneme depend on its local context, especially at the beginning and end of the phoneme.

One possible source of coarticulation consistent with the idea of local interactions is that the paths of the articulators under the control of a particular phoneme are partially determined by the positions of those articulators at the end of the commands for the previous phoneme in the sequence. In this case the different paths of the articulators are not planned (in the sense of being determined by the system ahead of time); rather, the articulator trajectories result from the interaction of the current articulator positions and the commands for the next phoneme. Lenneberg (1967), on the other hand, has proposed an explanation for coarticulation that suggests a much stronger planning component. He proposes that for some phonetic sequences XY, part of the signal to articulate Y must leave the brain *before* the signal to articulate X.

One way to begin to differentiate between these alternative explanations for coarticulation, as well as others that fall between these extremes, is to look at the control signals to articulators. Even with this information, however, there will be cases where an instance of coarticulation cannot be classified definitively as a result of planning or local interactions. For our purposes, however, a useful heuristic is that the longer the time between an interaction and the following element that engendered it, the more confidence we can have that this interaction represents the result of planning.

Prior to the work that we are about to examine, speech researchers had interpreted various coarticulatory phenomena as support for preplanning and as evidence for what the unit of planning might be (Kent and Minifie 1977 includes a good review of these ideas). An illustration of these phenomena can be experienced in normal speech. Some vowel sounds such as /u/ (as in *you*) require protrusion and rounding of the lips to be produced correctly. Most other vowels (such as the vowel /æ/ as in *bat*) cannot be produced correctly with the lips rounded. On the other hand, many consonants can be produced acceptably with or without lip rounding, although the presence or absence of lip rounding will affect how they sound. It has been observed that if a series of consonants, which are "neutral" as far as lip rounding is concerned, is followed by a vowel requiring lip rounding, then the consonants are also produced with lip rounding. For example, in the word *construe* the final vowel /u/ is produced with lip rounding. The rounding is also anticipated in the production of this word, being present on the /str/ sequence that begins the second syllable. To see this, compare pronouncing *construe* and *constrict*.

In the case of *construe*, the anticipation of the lip rounding spans three segments, phonemes, in one syllable. Anticipatory effects of lip rounding have been cited that span up to six consonants in a syllable (Kent and Minifie 1977), although such constructions are not possible in English. Similar effects have also been reported that cross syllable and word boundaries (Moll and Daniloff 1971), although these effects seem less common.

Demonstrations such as these of lip rounding certainly satisfy the requirement for evidence to support a claim of preplanning that a specific, motor-related aspect later in a sequence influences performance at some distance earlier in the sequence. Although these demonstrations are impressive in themselves, there do not appear to be many other examples in speech that exhibit these properties. This raises the possibility that these anticipatory effects might be due to some special mechanism rather than to general preplanning. Such effects are also limited in their scope: they have only been observed to span several segments within a syllable or across a single syllable/word boundary. If, for example, the syllable rather than the phoneme is the unit of sequence planning in speech, then most of these coarticulation effects would be interpreted, because of their limited range, as within-unit effects rather than effects of planning across units making up a sequence.

5.1.4 A Plan for the Rest of the Chapter

In the rest of this chapter we will look at experimental results that support several inferences about the nature of movement sequencing and motor programs. We will first examine regularities in movement-sequence performance in the domains of speech and typing. We will then look at a parsimonious model capable of describing these regularities, as well as at research that proposes and rejects other plausible alternative models. Finally, we will use the model to generate novel predictions about new data—predictions that, when borne out, will strengthen our confidence in the validity of the model.

5.2 Regularities Observed in Speech

The work we will be considering began with experiments by Monsell and Sternberg (1981). In their pioneering studies they asked practiced subjects to recite lists of words as quickly as possible. Most of their results related the durations of the utterances produced to the number of words in a list.

5.2.1 Method

Figure 5.1 outlines the procedure used by Monsell and Sternberg. Most of these experiments involved only a few (four or six) highly practiced subjects, who participated in a large number of trials over many days. The sequence to be spoken on a given trial was displayed on a CRT screen. The sequence consisted of a word or series of words, usually monosyllabic, to be spoken as a single fluent utterance (for example, *five, three, one, two* or *track, bay, rum*). The display remained on the screen for several seconds, and then the subject was allowed 2 or 3 seconds more to prepare for a "go" signal. To maximize preparation, the duration of this preparation period was fixed, and it ended with two signals, rhythmically spaced, 750 milliseconds apart. To discourage anticipations, there were occasional "catch" trials on which the "go" signal did not occur and subjects were not to respond; this occurred on 10 to 20 percent of all the trials.

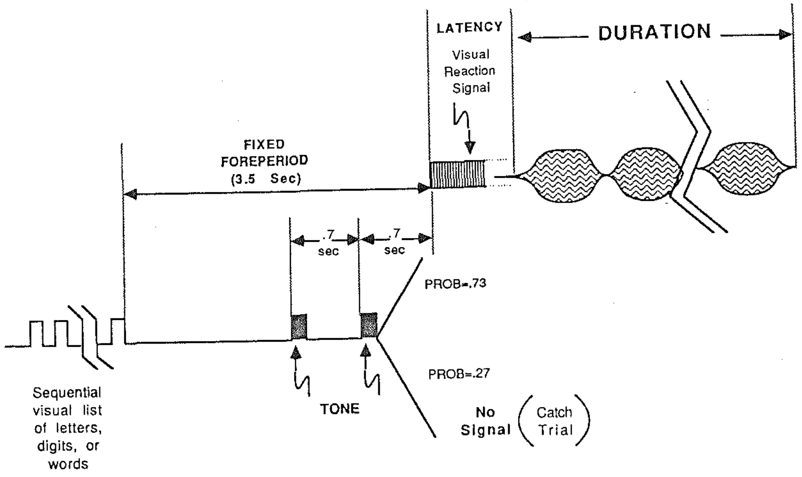


Figure 5.1
Procedure on one trial in a typical speech-production experiment. (After figure 1 in Sternberg, Knoll, Monsell, and Wright 1983; by permission.)

Many of these details, although important for an understanding of the conditions studied, appear not to be critical for obtaining the results of primary interest in this chapter. For example, in another condition studied by Monsell and Sternberg and also reported by Sternberg et al. (1980), the foreperiod varied randomly from 2.6 to 5.4 seconds with no warning signals or catch trials, but the overall pattern of the results remained the same. What may be critical, however, is that instructions and feedback encouraged subjects to produce the sequence (1) *correctly*, with the specified words in the correct order, (2) *fluently*, without stumbles or pauses, and (3) *with minimal time from start to finish*. To ensure (1) and (2), an experimenter constantly monitored all of the utterances and counted those as errors that were incorrect or insufficiently fluent; subjects were penalized for these errors and data from these trials were excluded from most analyses. To ensure (3), subjects were given numeric feedback after blocks of 15 to 20 trials indicating the average duration of their utterances and the number of errors. Scores computed from these values were compared to target scores, which were set at levels designed to improve on previous performance, and subjects were given monetary bonuses for beating their target scores.

5.2.2 Latency Data

When Monsell and Sternberg began these experiments, their initial goal was to examine how the time to start saying a list, the *latency*, would vary

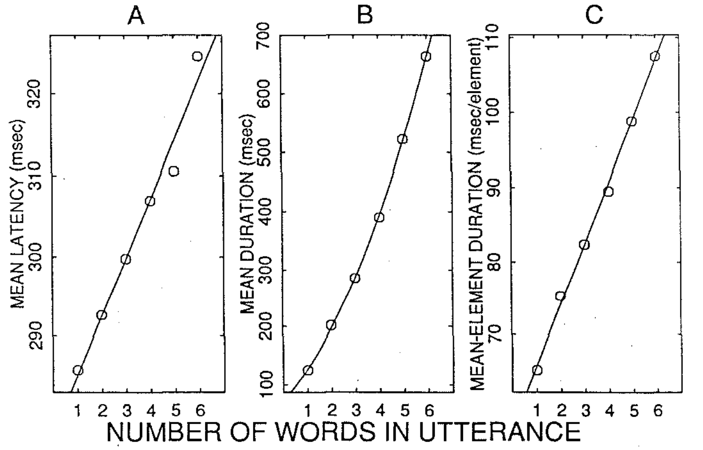


Figure 5.2
(a) Mean latency (b) utterance duration, and (c) estimated mean-element duration, as a function of the number of words for random sequences of digit names. The solid line represents a least-squares fit to the data in each case: for mean latency, $\hat{L}_n = 277 + 7.3n$; for mean duration, $\hat{D}_n = 40.9 + 55.0n + 8.6n^2$; and, for mean-element duration, $\hat{d}_n = 57.1 + 8.3n$. (Data taken from figures 3 and 5 in Sternberg, Knoll, Monsell, and Wright 1983; redrawn by permission.)

as a function of its length and composition. This interest stems from a long line of research based on the premise that, during the latency interval between the signal to respond and the beginning of the response, critical last-second computations are needed to construct the motor program for the upcoming movement (Henry and Rogers 1960). Much of this research has focused on the question of whether the latency to respond reflects the *complexity* of the upcoming movement (Hayes and Marteniuk 1976). In these experiments, changing the number of elements in the list was one way to vary the complexity of the material to be programmed.

Figure 5.2 shows data from one of Monsell and Sternberg's earliest experiments in which the utterances were made up of digits in random order. As figure 5.2a shows, Sternberg and Monsell found an effect of list length on the latency to begin reciting: mean latency increased approximately linearly with the number of words. The size of this *length* effect can be summarized by the slope of the fitted function in equation (1):

$$\hat{L}_n = \eta + \theta n. \tag{1}$$

Here n is the length of the list in words, $\hat{L}_n$ is the predicted latency for a list of length n, η is a constant intercept, and the slope parameter, θ, describes how the latency increases with list length.

5.2.3 Duration Data

Although these observations, in particular that of the functional form of this relationship, were important, Monsell and Sternberg also made another important observation. Along with the latency, the dependent variable in which they were primarily interested, Monsell and Sternberg measured the *duration* of the utterances: the time from when the subject began speaking until the utterance was complete. The results for duration are shown in figure 5.2b. These data show that mean duration, the average time to say the list of words, did not go up linearly with the list length, as one might expect, but rather increased approximately quadratically with the length of the list. This result is surprising since most intuitions about this performance suggest that durations should increase linearly with the list length.

To understand the implications of a quadratic duration function, consider the simpler, counterfactual case where the average duration of an item, in this experiment a digit, is roughly constant. We can represent this constant with the symbol d. In addition, we can use the symbol $\hat{D}_n$ to refer to the predicted duration of a list of length n. Using this notation, we expect the duration of a list of one word to be $\hat{D}_1 = k + d$, where k is a constant associated with the entire list. The value of k presumably is related to measurement error or "end effects," an issue we will discuss later. Similarly, we would expect the duration of a two-word list to be $\hat{D}_2 = k + 2d$, the duration of a three-word list to be $\hat{D}_3 = k + 3d$, and so on. We can generalize these equations for any value of n with the linear equation $\hat{D}_n = k + nd$.

As figure 5.2b shows, however, the actual durations that Monsell and Sternberg observed increased more than linearly with the number of items in the list. The upward curvature in these data is well described by the quadratic duration function in equation (2):

$$\hat{D}_n = \alpha + \beta n + \gamma n^2. \tag{2}$$

Here α, β, and γ are constants and n is, once again, the number of words in the list. The quadratic form of the mean duration function implies that the average duration of a single word in these utterances was not constant. If we assume that α represents measurement error or end effects on the entire list and correct for them, then the predicted average duration of an item in a list of length n, call this $\hat{d}_n$, is given in equation (3):

$$\hat{d}_n = \frac{D_n - \alpha}{n} = \beta + \gamma n \tag{3}$$

Equation (3) results by subtracting α from both sides of equation (2), since this is a value associated with the entire list rather than with an individual word, and dividing the result by n, since d_n represents the average duration of a single item obtained from the duration of a list made up of n words.[2] This algebraic manipulation makes it clear that, in Monsell and Sternberg's data, the average duration of *an individual word* rather than the *overall duration of the list* increases linearly with the length of the list.

Equation (3) suggests a useful form in which to examine the duration data from these experiments. Because the small degrees of curvature of the quadratic duration functions are hard to estimate and compare visually, often a set of observed utterance durations ($D_1, D_2, \ldots, D_n, \ldots$) is transformed into estimates of mean-element durations ($d_1, d_2, \ldots, d_n, \ldots$) using the formula in equation (3) so that the datasets can be compared more easily. To the extent that the utterance durations, D_n, are fit well by a quadratic function, the estimated mean-element durations, d_n, will be fit well by the linear function in equation (3). Figure 5.2c shows the mean-element transformation of the duration data and the best-fitting linear function.

5.2.4 Generality

The basic pattern of these results for latency and duration turns out to be robust. Sternberg et al. (1978) and Sternberg et al. (1980), for example, report similar patterns of data using utterances made up of

1. numbers in sequence like *two-three-four-five,*
2. weekdays in sequence like *Wednesday-Thursday-Friday,*
3. randomly ordered sets of weekdays like *Thursday-Monday-Tuesday-Saturday,*
4. reiterant sequences of weekdays like *Tuesday-Tuesday-Tuesday,*
5. randomly ordered lists of letter names,
6. novel and arbitrary lists of one- and two-syllable high-frequency nouns, and
7. lists of monosyllabic pseudowords like *vate-hane-vone.*

As we will see, a similar pattern of results for latency and duration is obtained for other complex, skilled movement sequences such as those produced in typewriting. In addition, Zingale and Kowler (1987) have observed very similar patterns of latencies and durations for sequences of

---

[2] In equation (3), D_n is the observed duration of lists of length n and the value of α is obtained from fitting the model in equation (2) to the duration data. Note also that the parameters β and γ are mathematically identical in equations (2) and (3). However, the estimates obtained for these parameters using standard (that is, unweighted) least-squares regression procedures to fit equation (2) to the utterance durations or equation (3) to the mean-element durations will not necessarily be exactly the same. This occurs because error in the observations at each value of n is weighted differently when fitting the two equations. An exploration of why this is so and what to do about it makes an interesting exercise.

saccadic eye movements. Finally, the effects occur not only for the highly practiced subjects studied by Monsell and Sternberg but also for unpracticed subjects, indicating that they are not a result of general practice or a lack thereof. The latency and duration effects also survive specific practice. If the same utterance is produced three times in a row, the overall durations decrease, but the form and slope of the latency and mean-element duration functions remain unaffected. Similarly, there are nonzero slopes for the latency and mean-element duration functions in typing even after 20 consecutive repetitions of the same string.

5.2.5 Additivity of Utterance Length and Word Length Effects

We turn now to a more detailed examination of another experiment by Monsell and Sternberg (1981; briefly reported by Sternberg et al. 1978). This experiment was designed to allow a comparison of the latency and duration function for lists of one- and two-syllable words. To reduce the effect of as many extraneous factors as possible, lists of one to four words were constructed in which all of the words were either one-syllable or two-syllable nouns with a high frequency of occurrence in written English (Kučera and Francis 1967). The two-syllable words were all ones normally produced with stress on the first syllable. More important, the first syllable of each of these words was one of the one-syllable words used in the experiment.[3] Examples of such embedded pairs include *bay* and *baby*, *cow* and *coward*, *rum* and *rumble*, *track* and *tractor*. With these stimuli, Monsell and Sternberg intended to create a manipulation as close as possible to the addition of an unstressed syllable to a given stressed syllable. One reason for trying to create such a precise manipulation was to ask whether the "unit" or "element" in terms of which these effects should be measured is the word, the syllable, or something else.

Figure 5.3 shows the latency, duration, and mean-element duration data from this experiment graphed with number of words as the independent variable. Figure 5.4 shows these same data, but this time the number of syllables is the independent variable. Before examining the details of these data, it is worthwhile simply to compare the two representations. It is striking that the representation in figure 5.3 results in fitted functions that are much more nearly parallel for the latency and mean-element duration data. The simplicity of representation in this and other similar comparisons convinced Sternberg et al. (1978) that the proper unit for the analysis for these data was not the syllable but probably something more similar to the word. Focusing first on the latency data, the slopes of the fitted functions for one- and two-syllable words are almost identical; the difference is 0.9 ± 1.1 milliseconds per word.[4]

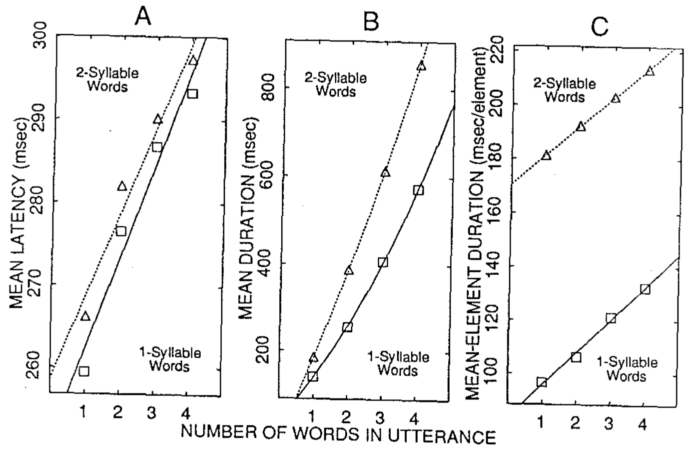


Figure 5.3
(a) Mean latency, (b) utterance duration, and (c) estimated mean-element duration, as a function of the number of words for utterances composed of one-syllable words (squares) and two-syllable words (triangles). The solid lines represent the least-squares fit for the one-syllable words: for mean latency, $\hat{L}_n = 251 + 11.1n$; for mean duration, $\hat{D}_n = 45.0 + 81.9n + 12.6n^2$; and, for, for mean-element duration, $\hat{d}_n = 83.3 + 12.3n$. The dotted lines represent the least-squares fit for the two-syllable words: for mean latency, $\hat{L}_n = 258 + 10.2n$; for mean duration, $\hat{D}_n = 2.8 + 171.8n + 10.5n^2$; and, for mean-element duration, $\hat{d}_n = 170.8 + 10.7n$. (Data taken from figure 15.3 in Sternberg, Monsell, Knoll, and Wright 1978 and figure 6 in Sternberg, Knoll, Monsell, and Wright 1983; redrawn by permission.)

As one might expect, despite its small effect on latency, the number of syllables in a word had a large effect on the duration function. This is best seen by looking at the mean-element durations in figure 5.3c. The intercepts of the fitted functions for one- and two-syllable words clearly differ: 83.3 milliseconds versus 170.8 milliseconds, respectively. An obvious interpretation of this is that the two-syllable words took, on average, 87.5 milliseconds longer to say. What is less obvious is that, although there was a large effect of the number of syllables on the *average* mean-element duration, number of syllables had little or no influence on the effect of list length or on mean-element duration. This is shown by the small difference

---

[3] The correspondence is not perfect, in some cases and for some dialects—a trained phonologist could discern and describe differences. For the purposes of this experiment, however, it appears to have been adequate.

[4] Here, as elsewhere in this chapter, the indication of variability, in this case 1.1 milliseconds per word, is the standard error of the mean based on the between-subject variability.

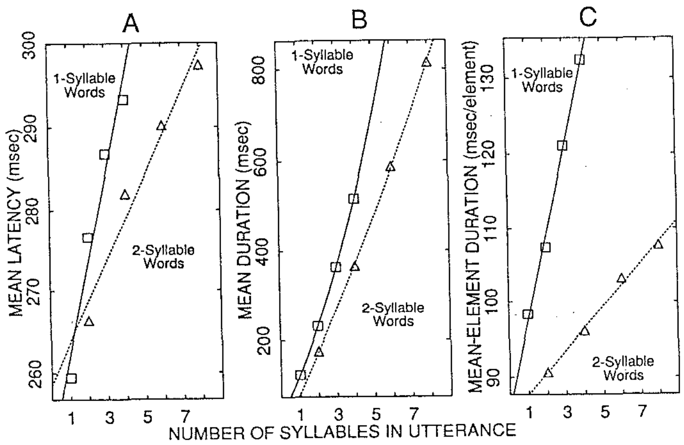


Figure 5.4
(a) Mean latency (b) utterance duration, and (c) estimated mean-element duration, as a function of the number of syllables for utterances composed of one-syllable words (squares) and two-syllable words (triangles). (Compare figure 5.3.) The solid lines represent the least-squares fit for the one-syllable words (these fits are identical to those in figure 5.3 for one-syllable words): for mean latency, $\hat{L}_n = 251 + 11.2n$; for mean duration, $\hat{D}_n = 45.0 + 81.9n + 12.6n^2$; and, for mean-element duration, $\hat{d}_n = 83.3 + 12.3n$. The dotted lines represent the least-squares fit for the two-syllable words: for mean latency, $\hat{L}_n = 258 + 5.1n$; for mean duration, $\hat{D}_n + 2.8 + 85.9n + 2.6n^2$; and, for mean-element duration, $\hat{d}_n = {}^*5.3 + 2.7n$. (Data taken from figure 15.3 in Sternberg, Monsell, Knoll, and Wright 1978 and figure 6 in Sternberg, Knoll, Monsell, and Wright 1983; redrawn by permission.)

between the slope estimates in these two conditions, 1.6 ± 1.2 milliseconds per word.

One summary of these results is that the effects of list length and number of syllables on mean-element duration and, perhaps to a lesser extent, on latency are additive; that is, the effects can be separated and their contributions added together to produce a good estimate of their combined effect. This is the simplest possible description for the combined effects of two factors on an observable variable. The fact that this is an adequate description in this case is a powerful observation with important theoretical implications that we will examine shortly.

5.2.6 Summary

Before we turn from speech to similar phenomena observed in typewriting, it may be useful to summarize the points made so far:

1. The latency to begin an utterance increases by a fixed amount θ for each element in the utterance.
2. The mean-element duration for an utterance increases by a fixed amount γ for each element in the utterance.
3. Although this is not documented systematically in the preceding discussion, the estimates of θ and γ tend to be quite similar.
4. The effects of number of elements and element size on latency are additive.
5. The effects of number of elements and element size on mean-element duration are additive.

5.3 Regularities Observed in Typewriting

5.3.1 Latency and Duration Data

Sternberg et al. (1978) have demonstrated in a number of experiments that *burst* typing of nonsense materials exhibits effects of list length on both the latency and the duration of the sequence. In burst typing, the subject is required to type a short sequence (one to six) of previously presented letters (usually all consonants) as quickly as possible. This task can be distinguished from *transcription* typing, in which a longer text is copied from another source. The burst typing paradigm has the advantage, for this work, that it has little or no perceptual (reading) component.

Figure 5.5 shows data from a typical typing experiment. The similarity of many aspects of these data to the speech data suggests that the phenomena in both modalities are the result of a general mechanism or strategy of motor control. Note that figure 5.5 also contains a comparison of between-hands and within-hand typing sequences. This manipulation is very similar in spirit to the one- versus two-syllable manipulation for the speech experiments. It is well known that the time between successive keypresses is longer if the keys involved are typed by the same hand rather than by different hands. Thus, for instance, the average time between keypresses for two letters typed by the same hand (for example, *J* and *K* or *A* and *D*) might be 200 milliseconds for a skilled typist, even though these keys are usually struck by different fingers. For comparison, the average time between two keypresses for letters typed by different hands (for example, *J* and *A* or *D* and *K*) might be 120 milliseconds for the same typist. As figure 5.5c shows, this difference, like the effect of one- versus two-syllable words, is reflected in the intercept, β, of the function relating mean inter-keypress time to the number of keypresses (equivalently, the linear parameter of the overall duration function), not the slope parameter, γ (the quadratic parameter of the overall duration function). Thus, once again there is clear additivity between the effect of sequence length and an element-size factor, in this case same-hand versus alternating-hands transitions, that has a large effect on the time to produce a single element.

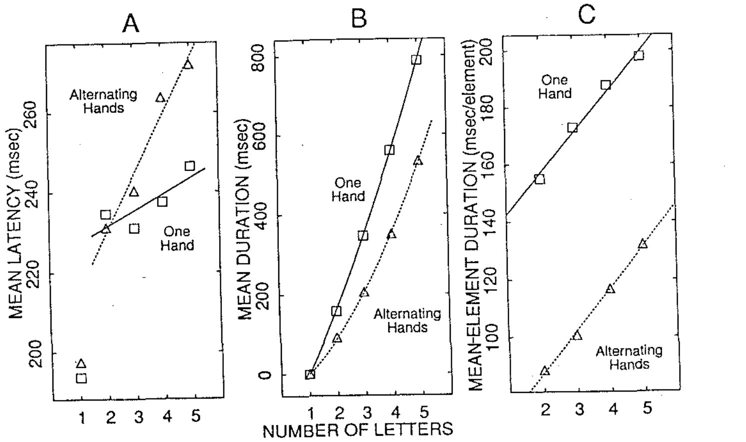


Figure 5.5
(a) Mean latency, (b) duration, and (c) estimated mean-element duration as a function of the number of letters in typewritten sequences of one to five letters. Sequences were typed either using all one hand (squares) or using a strictly alternating sequence of hands (triangles). The solid lines represent the least-squares fit for the one-hand condition: for mean latency (fit from $n = 2$ to $n = 5$ only), $\hat{L}_n = 223 + 4.1n$; for mean duration (constrained to pass through zero when $n = 1$), $\hat{D}_n = 142.9n + 14.3n^2$; and, for mean-element duration, $\hat{d}_n = 142.9 + 14.1n$. The dotted lines represent the least-squares fit for the alternating-hand condition: for mean latency (fit from $n = 2$ to $n = 5$ only), $\hat{L}_n = 200 + 14.9n$; for mean duration (constrained to pass through zero when $n = 1$), $\hat{D}_n = 72.7n + 14.7n^2$; and, for mean-element duration, $\hat{D}_n = 71.9 + 15.2n$. (Data taken from figures 15.5 and 15.6 in Sternberg, Monsell, Knoll, and Wright 1978; redrawn by permission.)

5.3.2 Inter-Keypress Time Data

Throughout this discussion of the effect of length on duration, we have been considering mean-element (word or inter-keypress) durations without ever considering the element durations themselves. An examination of speech or typing at this level requires measuring the duration of each element (word or inter-keypress time) as a function of the list length and the serial position of the item within a list. Although these measurements are hard to make for speech, they are quite straightforward to make for typewriting. The primary data collected for typing are the times and identities of a sequence of key-closures. From these data, it is simple to compute the intervals between successive keypresses.

Figure 5.6, a reanalysis of the data summarized in figure 5.5, shows typical inter-keypress time data for sequences made up of all same-hand transitions in one case and all alternating-hands transitions in the other. In graphing these data, Sternberg et al. (1978) chose to line up points for sequences of different lengths relative to the last keypress in each sequence. An alternative would have been to graph these data so that the first keypress in each sequence lined up. A priori there is no good way to choose between these and any of a number of other ways to establish (at least implicitly) correspondences between the various serial positions at each list length. (Are initial keypresses in sequences of different lengths necessarily more similar than the last keypresses in those same sequences?)

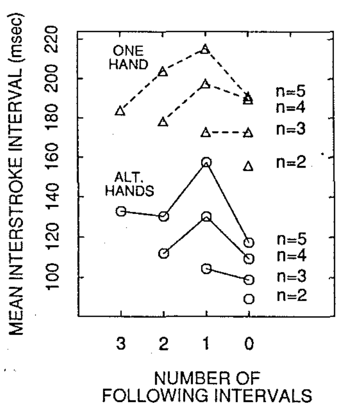


Figure 5.6
Data from figure 5.5 reanalyzed to show the time intervals between successive keypresses. Open triangles mark the data from the one-hand condition. Open circles mark the data from the alternating-hands condition. (Data taken from figure 15.6 in Sternberg, Monsell, Knoll, and Wright 1978; redrawn by permission.)

Two aspects of the data in figure 5.6 merit particular attention. First, the effect of sequence length is distributed fairly evenly across all of the keypresses in the sequence. To see this, consider first as an example the last inter-keypress times in the sequence for each length. These are represented by the points lined up vertically at the right side of the graph. Even though all of the other keypresses have been made, the time from the next-to-last keypress to the last keypress is longer when $n = 3$ than when $n = 2$, when $n = 4$ than when $n = 3$, and so on. This pattern of these data points is referred to as *dominance*. Looking at the other serial positions (vertical columns) in figure 5.6, it is clear that dominance holds for all of them. It is also of interest to consider the first keypress in each sequence. The data points representing these keypresses lie along a rough diagonal from the top-left to the bottom-right for the two sets of data. Again we find dominance according to list length; the time from the first to the second keypress increases as the number of keypresses to follow increases. Thus, even the interval between the first and second keypress in a sequence reflects the length of the entire sequence.

The overall increase in inter-keypress time with list length is not, however, the only factor determining the inter-keypress times in figure 5.6. A second interesting point about these data is that there are clearly strong effects of serial position within a list. The last keypress in lists of different lengths is one of the fastest, the second-to-last keypress in each list is always the slowest, and so on. These serial position functions cannot be explained by some simple artifact in the design of this experiment: Sternberg et al. (1978) were careful to ensure not only that each key occurred equally often at each serial position for each list length but also that the distribution of transitions between keypresses was consistent within conditions across serial position and list length. This makes it all the more interesting that, despite the apparent complexity of these serial position functions, we know from figure 5.5c that averages of these data across serial position lie on straight lines as a function of list length. Further, although the detailed shapes of the serial position functions differ somewhat for the alternating-hands and same-hand data, averaging across serial position for these two conditions results in parallel lines as a function of list length.

Sternberg et al. (1980) have reported similar serial position functions for the durations of spoken words within lists. As noted above, these measurements are quite difficult to make. The measurement process also appears necessarily to involve some degree of subjective human judgment or the imposition of largely arbitrary criteria. For these two reasons we will not dwell on them further here. The important point is that although there are systematic differences between the patterns of element-duration functions for speech and typing, dominance as a function of list length holds in both modalities. Put another way, in both cases there is a general slowing of all items in a list as a function of list length. This then is a case where a global property of a sequence affects performance at every point within the sequence. This is exactly the pattern of data that is taken as strong evidence for preplanning.

In addition, however, both domains reveal a complexity of data at the level of individual element durations that is absent when these durations are averaged to produce mean-element durations. This additional complexity in the lower-level measurements can be taken as evidence that the mean element is the best or most appropriate level for studying the mechanism underlying the regularities of the latency and duration data. Of course, this conclusion may be wrong, but it certainly defines an obvious starting point for model building.

5.4 Understanding the Regularities

So far we have concentrated on outlining a set of performance regularities in speech and typing, some of which were decidedly unexpected when they were first observed. This research required subjects to produce short action sequences, in a context that provided both the incentive and the opportunity to use advance planning. In this situation the sequence duration increased as a quadratic function of the number of elements in the sequence. This corresponds to a linear increase in the average element duration, an increase that is distributed over all of the elements in the sequence. Thus, a characteristic of the whole sequence influences the execution of each of its elements, precisely the type of evidence that suggests that a representation of the whole sequence—a motor program—exists before the sequence begins and is used to control the production of the sequence. If sequence length is measured as words for speech or keypresses for typing, then the size of the length effect is the same for words with different syllabic structure and duration and for keypress sequences made at different rates because of different hand combinations: the resulting mean-element duration functions are parallel and vertically displaced. This additive invariance suggests that these measures of sequence length may be theoretically significant.

5.4.1 The Subprogram Retrieval Model

Figure 5.7 sketches, in flowchart form, a model to explain the performance regularities just described. This model was first outlined by Sternberg et al. (1978) and has subsequently been restated and refined a number of times (Monsell and Sternberg 1981; Sternberg et al. 1983; Monsell 1986). The basis for this model is an attempt to describe one way a motor program might be used. The model incorporates a number of assumptions.

1. During the preparation interval the speaker or typist constructs a *program*, made up of *subprograms*, specifying the elements of the utterance to be spoken or the string to be typed and the sequence of those elements. Although we will not look at the evidence for this (see Monsell and Sternberg 1981 or Monsell 1986), this program is believed to be stored in a motor-program buffer distinct from short-term or working memory. Since little knowledge is currently available about the level of abstraction used for representing this information, the model is silent on this issue.
2. When the "go" signal is detected, the initial unit or subprogram is *retrieved* and then a *command* process initiates activity, ultimately resulting in the pattern of activity that is appropriate to generate the element specified by the subprogram. For speech, this would be vocal

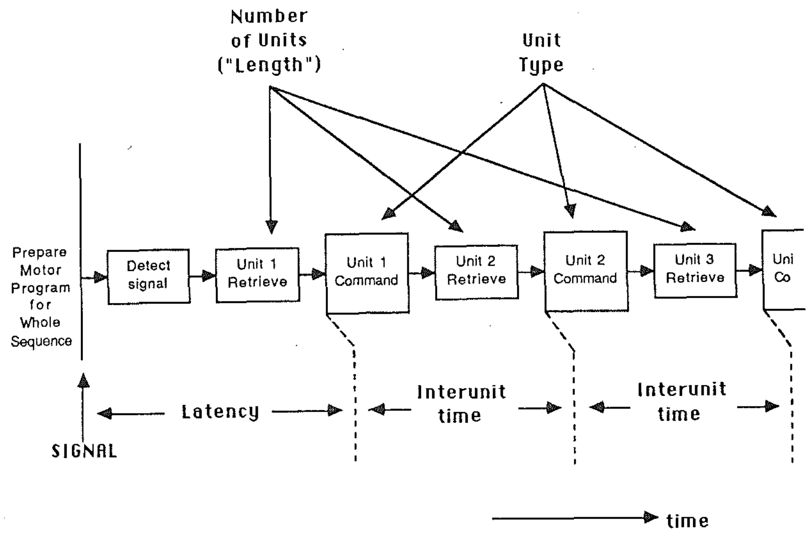


Figure 5.7
Subprogram retrieval model. (Redrawn, by permission, from figure 5 in Sternberg, Knoll, Monsell, and Wright 1989.)

activity; for typewriting, this would be hand and finger activity. The cycle of retrieval followed by command stages repeats until every element in the sequence has been executed.

3. The duration of the retrieval stage is a function of the number of subprograms that make up the program. The more subprograms that are simultaneously in a state of preparation, the more time it takes to "activate," "locate," "retrieve," or "select" each one. (These terms are all listed because any might supply an appropriate metaphor for this process. This process is meant to involve nothing more than gaining access to or passing control to the appropriate subprogram.)

4. The duration of the retrieval stage is influenced not by the composition of the subprograms but only by the number of subprograms. It is not known why the retrieval duration is a linear function of the number of subprograms. One possible mechanism is sequential *search* through a set of subprogram "directory entries" or "addresses." Under that proposal, linearity of retrieval time follows from simple properties of the search process. This mechanism can also be elaborated (see question 5.1) to explain the simultaneous simplicity of the mean-element duration functions and the complexity of element duration as a function of serial position. However, other mechanisms could undoubtedly account for these findings and as yet no compelling evidence has been found to support this particular proposal.

5. The duration of the command stage depends only on the composition of the subprogram being processed and not on the number or composition of the other subprograms, if any, in the sequence. This amounts to a claim that the actions of the command stage are purely local, made without reference to any information contained in the rest of the program.

6. The command and retrieval stages are "processing stages" in the strict sense defined by the assumptions of Sternberg's (1969) "additive factors" method. When subjects are required to complete the sequence as quickly as possible (and perhaps only under those conditions), the duration of a subprogram is the sum of the durations of its retrieval stage and its command stage.

5.4.2 Relating Model Stages and Equation Parameters

The structure of this model allows us to map various of its processes to the parameters in the fitted functions describing the latency and duration of sequences, equations (1)–(3). Examination of figure 5.7 may help to make these relations clearer.

Equation (1), describing the latency to begin a sequence, incorporates two parameters: the intercept parameter η and the slope parameter θ. The model posits two processes operating during the latency interval before the effects of the first command stage are reflected in output activity. The signal-detection process is posited to be independent of sequence length and element size. The time for signal detection is reflected in the intercept parameter, η, in equation (1). The latency interval also includes the time for the retrieval of the first subprogram. This time, according to the model, depends on the length of the sequence and is represented in equation (1) by the term θn.

Inevitably, the mapping of terms in the model to the elements in equation (1) is not completely straightforward. The intercept parameter η will additionally reflect a delay, which could be as long as 100 to 200 milliseconds, between the initiation of commands during the command stage for the first subprogram and the initiation of peripheral motor activity for the first element. The duration of this delay is assumed to be independent of both sequence length and element size. Unlike the time to detect the signal, this delay presumably would also not be changed by changes in the detectability of the "go" signal or the probability of catch trials. Of more theoretical importance, the latency interval may also include a chunk of the activity associated with the first command stage, suggesting that there could be an effect of element size on the latency. In speech, this contamination occurs because it is necessary to discriminate the onset of the speech signal from background noise; this process inevitably misses a little of the speech. The problem is probably worse for typing. Here, the keypress that terminates the latency interval occurs well into the movement trajectory that is controlled by the first command stage. These, or other similar effects, may be sufficient to account for effects of element size on latencies.

The processes posited by the model are also reflected in the parameters of the duration function. The duration of each element spans time taken by one command stage and one retrieval stage. Considering the mean-element duration function, equation (3), the intercept parameter, β, reflects the changes in command-stage duration associated with changes in the element size. Similarly, the term γn reflects how mean-element duration varies when retrieval time increases for sequences with more subprograms.

As discussed previously, the parameters β and γ have the same interpretations for the overall duration function, equation (2). The intercept parameter α in that equation reflects several factors that occur only once in the production of a list. For instance, it is likely that the criteria used to identify the beginning of the first element or the end of the last element in the external behavior systematically deviate from the points controlled by the beginning and end of the first and last command stages, respectively. This will lead to a systematic under- or overestimate of the duration of the performance, and this measurement error will be reflected in α. Similarly, any tendency of the subject to produce the first or last element in a sequence so that its duration differs systematically from what it would have been at internal positions in the sequence will be reflected in α.[5]

This model, along with the mapping of its elements to parameters in the fitted functions, describes all of the regularities we have discussed so far.

5.5 Predictions of the Model and Their Assessment

The descriptive success of the proposed model allows us to organize and make sense of a large body of data generated from two different response modalities. This should not be surprising, since the model was created to explain these data. A second approach to exploring a model such as this is to seek to confirm or disconfirm predictions made by the model that might

---

[5] One example of this type of phenomenon from speech is phrase-final lengthening (Vaissiere 1983). This describes the strong tendency most speakers have to lengthen a syllable when it is the last syllable in a phrase, in some cases doubling the duration of the same syllable in a neutral context. This lengthening is often interpreted as a prosodic signal of the syntactic phrase structure of an utterance. This lengthening is also present under the conditions used in the experiments reported here, despite instructions that discouraged it, a reward system that penalized it, and the lack of a communicative purpose for it.

otherwise be unexpected or counterintuitive. We will explore two such predictions.

5.5.1 The Connection between Latencies and Mean-Element Durations

The model posits that two terms in the fitted functions, θn for the latency function and γn for the duration function, depend solely on the time required to retrieve one subprogram. Since there is no compelling reason to suspect that the value of n changes between the latency interval and the duration interval, the model implies that these two terms should have similar values and thus that $\theta = \gamma$.

Over the course of the many experiments done to explore the regularities captured by this model, only a few of which are described here, a large range of values for θ and γ have been estimated for many subjects under a large variety of conditions. These parameter estimates have ranged from low values, 3 or 4 milliseconds per element, to much higher values, on the order of 30 milliseconds per element. Since most of these experiments are not described here, it would be inappropriate to make a detailed comparison of the values of θ and γ across these experiments. However, it can be said that the correspondence is remarkably good, supporting this claim of the model.

5.5.2 Intermittency of the Effect of Length on Sequence Production

The sequences of activity controlled by the model are, in both speech and typing, relatively continuous. In typing, for instance, although the keyboard registers that a key has been pressed at one discrete instant when the key travels past a designated point on its path, the finger and wrist movements that cause the key to move are smooth. Similarly for speech: although we have the perception of hearing separate words in the speech stream, if we were to look at any of several graphical representations of the energy in speech, it would often be hard to decide where one word stopped and the next started. Against this background, the cyclic progression of nonoverlapping stages posited by the model seems distinctly unnatural.

According to the model, the time between the initiation of one element and the initiation of the next is filled first by a command stage and then by a retrieval stage. Of these two stages, the duration of only one, the retrieval stage, is lengthened as the length of the list increases. Taken together, these facts suggest that the influence of sequence length on the output activity should be *intermittent*.

Consider what this suggests about finger movements in typing or vocal articulator movements in speech. A command stage initiates and largely controls the trajectory for one element of the output sequence according to the directions contained in the most recently retrieved subprogram. Because the command process is not influenced by sequence length, it is possible that the corresponding portion of the trajectory for the element produced by this subprogram might also be independent of sequence length. This possibility depends on there being a strong moment-to-moment coupling between the command/retrieval processes and the observed trajectories. This coupling cannot be perfect, however, since the movements of speech and typing do not stop abruptly during the interval corresponding to the next retrieval process (it is effects of sequence length that the model suggests are intermittent, not the movement itself). One possibility is that activity initiated during the command process is carried forward under the control of lower levels in the motor system during the subsequent retrieval interval, a time when no further commands are being issued. If this is all approximately correct, then the extra time required for the retrieval process with longer sequence lengths should be *localized* within the production of each element of the sequence. To go even farther out on this limb, intuitively we might expect the localized effects of length to occur near the end of each unit, when the motor-sequence controller would need to retrieve the next subprogram.

The most interesting alternative to the prediction of localized effects is that the length effect is *distributed* throughout the duration of each output element. The empirical question has many similarities to that underlying the examination of the element durations as a function of serial position and sequence length. In both cases the primary question is whether the effect of length is localized at specific points or distributed across the entire activity. One difference between these analyses is that the model predicts the effects of length will be distributed across the elements in the sequence but that the effects of length will be localized within the segments of those elements.

This prediction seems counterintuitive and implausible. The measurements needed to confirm or deny this hypothesis are detailed and require either new analysis techniques or new instrumentation. Sternberg and his colleagues have, however, looked for evidence to confirm or disconfirm this prediction both in speech (Sternberg et al. 1980, 1983, 1989) and in typing (Sternberg et al. 1983); surprisingly, exactly the predicted pattern of results was found in both cases. Here we will consider only the procedure and results for the speech experiment.

To look for localization of the effects of sequence length in speech, Sternberg et al. (1980, 1989) had subjects produce sequences containing from one to five two-syllable words. The words themselves were chosen in an attempt to simplify the segmentation process. Both syllables of each word nominally began with a closure; all the words normally had stress on the first syllable (for instance, *copper* and *token*). Figure 5.8 shows the idealized energy envelope for one of these words. Using a sophisticated segmentation algorithm, which was developed for this research based on techniques originally used for speech recognition, each word was decomposed into six segments roughly corresponding to the phonetic categories of consonant-closure, consonant-release, and vowel for each syllable.

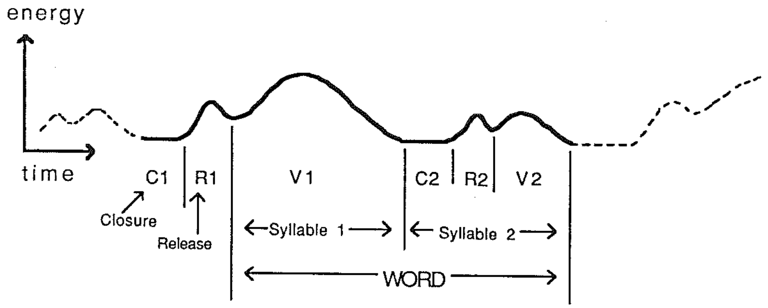


Figure 5.8
Schematic demonstration of the segmentation of a two-syllable spoken word (for instance, *copper* or *token*) into six segments based on an idealized energy envelope. (Redrawn, by permission, from figure 9 in Sternberg, Knoll, Monsell, and Wright 1989.)

The upper panel of figure 5.9 shows the mean-element duration function, which is similar to those found previously. The slope in this case is about 11 milliseconds per word. The bottom panel shows duration as a function of length for each of the six segments identified in these words. Note that these segment durations add up to the mean-element durations in the top panel. Not surprisingly, the longest segment is the stressed initial vowel (V1). Most of the effect of length is localized in the vowel of the unstressed second syllable (V2) even though its duration is, on average, only about 40 percent of that of V1. There are also marginal increases in length for the durations of the first closure (C1) and the first release (R1). Because the second syllable of one word is followed by the first syllable of the next word, these three segments are contiguous. This surprising observation of localization in the effects of sequence length fits with the predictions outlined earlier about the peripheral manifestations of intermittency in the cyclic control process posited by the model to control the production of these sequences.

5.5.3 Limits of the Model's Applicability

Although the generality and predictive success of the model are encouraging, it is important not to overgeneralize from it. For example, it would be wrong to expect that any response modality will exhibit the complete set

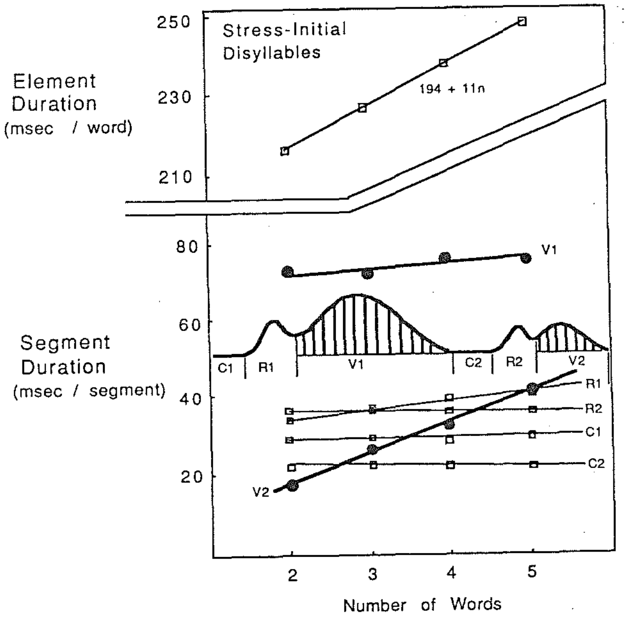


Figure 5.9
Mean-element duration and mean-segment duration of two-syllable spoken words. The upper panel shows the mean-element duration data. The line represents the least-squares: $\hat{d}_n = 194 + 11.1n$. The lower panel shows the decomposition of the mean-element durations into six mean-segment durations. (Redrawn, by permission, from figure 10 in Sternberg, Knoll, Monsell, and Wright 1989.)

of regularities outlined in this chapter. A counterexample to this generalization can be found in the work of Hulstijn and Van Galen (1983). They have reported that subjects rapidly producing prepared sequences of handwritten characters show a latency effect with increased sequence length but not a rate effect. This dissociation of the latency and rate effects might be troublesome for the model if handwriting were a domain in which movement timing is a free parameter. However, it has been shown that the control system for handwriting keeps stroke duration relatively constant despite changes in other factors such as writing size and pen-to-writing-surface friction that might be expected to cause changes in stroke duration (Denier van der Gon and Thuring 1965; Hollerbach 1981). Thus, even if a cycle of command and retrieval processes underlies the production of handwritten sequences, a rate effect would not be observed unless those processes limit the rate of production. An analogy has been suggested by Monsell: The rate at which a highway is constructed can be limited either by the efficiency of the bureaucrats responsible for obtaining permissions, purchasing land, and so forth, or by the work rate of the construction crews. Only in the former case would measurements of the rate of highway construction tell us anything about the performance of the bureaucracy (the planners).

Suggestions for Further Reading

The timing results for speech and typing described here are only selected highlights from a larger set of experiments undertaken to explore the properties of motor programs in these domains. Discussions of many of the important issues omitted can be found in the works cited. Among these topics is a more precise determination of the *unit* of programming for speech and typing. It turns out, after a more thorough investigation, that both the word in speech and the keypress in typing are only simplified approximations to the actual units in those domains. In particular in typing, immediate repetitions of a single character, *doublets*, appear to form a single unit. Sternberg and Knoll (unpublished research described briefly in Sternberg et al. 1983) test this hypothesis and use it to construct an ingenious test of the overall model by embedding single doublets in larger strings and then examining the implication of their presence for the effects on length of mean-element duration. It is also possible to explore, as Monsell (1986) has done, what the ultimate capacity of these programs is, how and where they are stored, and how they are maintained.

The findings described here might be explained from other perspectives or modeled using different techniques. Although a number of alternatives to explain pieces of the patterns described here have been proposed, tested, and, in many cases, rejected (see, in particular, Sternberg et al. 1978, 1980), no alternative proposals have been suggested to explain the broad range of these results. One alternative approach that has been explored for typing is that of using parallel activation-triggered schemas to control movement sequencing and trajectories (Rumelhart and Norman 1982; Norman and Rumelhart 1983). This particular simulation was developed to explain a different set of phenomena observed in transcription typing. Although this model has many appealing aspects and could certainly be usefully extended to cover other domains such as speech, it is not clear how it could be extended to explain the regularities described here.

A second approach that appears promising as a way to explain the data described here is the hierarchical editor model described by Rosenbaum and his colleagues (Rosenbaum 1985; Rosenbaum, Inhoff, and Gordon 1984; Rosenbaum, Hindorff, and Munro 1987). Using a model that represents movement sequences as hierarchical tree structures, they are able to explain, at least qualitatively, many of the results reported here for latency, duration, and element durations. In addition, this model provides a rational treatment of many often seemingly mysterious phenomena that occur when a choice must be made at the last instant between different but related movement sequences.

A third alternative is the class of explanations based on the notion of general-purpose processing capacity. Appeals to capacity-oriented explanations are not unusual in cognitive science. One major problem with them is that the notion of capacity is often poorly defined; almost any pattern of results can be interpreted post hoc in terms of heavy capacity demands here, light ones there, and so on. One experiment done originally by Monsell and Sternberg (1981) and repeated on a larger scale by Monsell (1986) seems to exclude most versions of this model as possible explanations for the results discussed here. In the more recent of these experiments, subjects with memory spans between 7.2 and 8.3 prepared and produced sequences of one to five ordered weekday names. In addition, during the production of the sequence, they were required to remember lists of zero, one, three, or five digits. Surprisingly, although Monsell observed the normal effects of sequence length on latency and duration, there was little or no effect of memory-load size on the latency or duration data; this despite the fact that the combined lists were well above the subjects' memory span in some conditions. This is not to say that the added load did not make the overall task harder (requiring more capacity). It is not easy to maintain a concurrent memory load, but doing so does not seem to interfere with the preparation or production of rapid utterances as we might expect from a general-purpose processing capacity account.

Questions

5.1 One possible mechanism to instantiate the retrieval stage described in section 5.4.1 is the process of search through a partially ordered list. The challenge in theorizing about the retrieval stage is to conceive of a process that simultaneously predicts the linearity of the mean-element duration function and the complexity of the interresponse times as a function of list length and position within the list. An appropriately defined search process can satisfy both of these requirements.

Consider a mechanism in which information identifying each item in a sequence of items to be performed (and only those items) is stored in a buffer. Once one item in the sequence has been produced, it is necessary to search this buffer for the information about the next item. We might assume that this search process is (1) *serial* (items in the buffer are considered one at a time), (2) *self-terminating* (the search process stops as soon as the next item is identified), (3) *minimal* (each item in the buffer is examined only once in a particular search), (4) *fixed-time* (each decision takes, on the average, the same amount of time), and (5) *unordered* (the order in which the items in the buffer are checked is random related to their order in the sequence to be produced). A search process defined this way would produce the required linear mean-element duration function. Why is this so? How is the comparison time of the search process related to the parameter γ described in equations (2) and (3)?

This search process does not produce the correct results for element duration as a function of serial position and list length. What functions does it produce instead? Now consider what happens if we relax the constraint that the search process be unordered. How might the search process now generate both the linear mean-element duration function and the correct predictions for mean-element duration as a function of serial position and length?

5.2 The model proposed in this chapter and the data presented to support it deal with planning movement sequences at the highest level. What if you wanted to look for evidence of sequence planning based on trajectories of your hand moving in space? What would you look for in these trajectories as evidence for planning? How would you distinguish planning effects from unplanned, local interactions? Do you have intuitions about what it means to "plan," other than those discussed in this chapter, that might be useful in this analysis?

5.3 Many theorists are intrigued by the possibility that motor programs are organized hierarchically (see, for example, Greene 1972; Rosenbaum 1985; Rosenbaum, Inhoff, and Gordon 1984; Rosenbaum, Hindorff, and Munro 1987). How might you distinguish a hierarchically organized motor program from one having a simple linear structure?

References

Denier van der Gon, J. J., and J. Ph. Thuring (1965). The guiding of human writing movements. *Kybernetik* 2, 145–148.

Greene, P. H. (1972). Problems of organization of motor systems. In R. Rosen and F. M. Snell, eds., *Progress in theoretical biology*, vol. 2. New York: Academic Press.

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

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

Hollerbach, J. M. (1981). An oscillation theory of handwriting. *Biological Cybernetics* 39, 139–156.

Hulstijn, W., and G. P. Van Galen (1983). Programming in handwriting: Reaction and movement time as a function of sequence length. *Acta Psychologica* 54, 23–49.

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

Kučera, H., and W. N. Francis (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

Lenneberg, E. (1967). *The biological foundations of language*. New York: Wiley.

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

Moll, K. L., and R. G. Daniloff (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* 50, 678–684.

Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In H. Heuer and C. Fromm, eds., *Generation and modulation of action patterns*. Berlin: Springer-Verlag.

Monsell, S., and S. Sternberg (1981). Speech programming: A critical review, a new experimental approach, and a model of the timing of rapid utterances: Part 1. Unpublished technical memorandum, AT&T Bell Laboratories. [Available from the authors.]

Norman, D. A., and D. E. Rumelhart (1983). Studies of typing from the LNR research group. In W. E. Cooper, ed., *Cognitive aspects of skilled typewriting*. Berlin: Springer-Verlag.

Rosenbaum, D. A. (1985). Motor programming: A review and scheduling theory. In

H. Heuer, U. Kleinbeck, and K.-H. Schmidt, eds., *Motor behavior: Programming, control, and acquisition*. Berlin: Springer-Verlag.

Rosenbaum, D. A., V. Hindorff, and E. M. Munro (1987). Scheduling and programming of rapid finger sequences: Tests and elaborations of the hierarchical editor model. *Journal of Experimental Psychology: Human Perception and Performance* 13, 193–203.

Rosenbaum, D. A., A. W. Inhoff, and A. M. Gordon (1984). Choosing between movement sequences: A hierarchical editor mode. *Journal of Experimental Psychology: General* 113, 373–393.

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

Saltzman, E. (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology* 20, 91–163.

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

Sternberg, S., R. L. Knoll, S. Monsell, and C. E. Wright (1983). Control of rapid action sequences in speech and typing. Invited address given by Saul Sternberg to the 1983 Annual Meeting of the American Psychological Association. [Text available from the first author.]

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

Sternberg, S., S. Monsell, R. L. Knoll, and C. E. Wright (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*. New York: Academic Press.

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

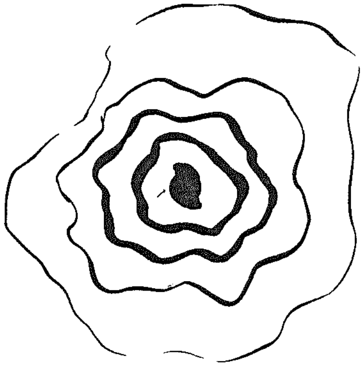
Vaissiere, J. (1983). Language-independent prosodic features. In A. Cutler and D. R. Ladd, eds., *Prosody: Models and measurements*. New York: Springer-Verlag.

Zingale, C. M., and E. Kowler (1987). Planning sequences of saccades. *Vision Research* 27, 1327–1341.

Chapter 6

Action and Free Will

Alvin Goldman



6.1 Two Images of the Behaving Organism

There are two ways of viewing human actors. The first is amply illustrated by the preceding chapters, which deal with locomotion, reaching and prehension, oculomotor control, and so on. As these chapters indicate, the activities in question are outputs of a highly complex physical system, ultimately to be understood in terms of its musculature and neurocircuitry. At a suitable level of analysis, persons are just physicochemical systems whose neurons are subject to the same electrochemical laws that govern wholly nonbiological systems. Although science has not yet identified all the relevant underlying principles, there *are* orderly, lawful patterns governing human behavior. Thus, the scientific image of human behavior is just a special segment of the general science of physical systems.

In at least apparent contrast to this scientific image of human beings is a second image: that of persons as freely choosing agents. Suppose you are deciding whether to go to a play tonight, and you weigh the pros and cons. The show is almost sold out, so you may get a poor seat; indeed, you