Control of Serial Order in Rapidly Spoken Syllable Sequences

PETER C. GORDON

Harvard University

AND

DAVID E. MEYER

University of Michigan

Serial-order information for speech production may be characterized in terms of three alternative classes of models: element-to-element coding, element-to-position coding, and hierarchical coding. Three experiments were conducted to test canonical versions of these models as accounts of the representation of syllable order in short utterances. A response priming procedure was used in which subjects prepared to say a "primary" utterance containing four syllables, but sometimes switched and said a "secondary" utterance containing the same syllables in a different order. Performance (reaction time, duration, and error rate) on the secondary utterance was measured as a function of the ordering of its syllables relative to those of the primary utterance. Experiment 1 supported a hierarchical coding model consisting of a tree with binary branches. In particular, subjects produced the secondary utterances faster when they were compatible with a binary-tree representation of the primary utterances than when they were not. Experiment 2 supported the hierarchical coding model over an element-to-element coding model, and Experiment 3 supported it over an element-to-position model. The binary structure of the hierarchical coding of syllable order may be based in the rhythmic patterns of speech. © 1987 Academic Press, Inc.

Speech production is a translation from the abstract to the concrete; an idea must be elaborated into a set of articulatory movements. One critical component of this elaboration is the selection of linguistic units. These units provide an interface between the realm of meaning and the realm of articulation. A second critical component is the ordering of linguistic units. The creative nature of language rests on the ability to combine a finite set of linguistic units in infinite ways (Chomsky, 1963). This article focuses on one aspect of the ordering process: the manner in which syllable order is represented in fully prepared utterance plans.

Language can be described at many levels, ranging from sentences and words to phonetic segments and articulatory gestures. Traditionally, the relations between the units at these different levels are given a hierarchical description, such that higher level units (e.g., words) specify lower level units (e.g., phonemes). Such descriptions are consistent with our intuitions about language structure, relating to regularities

within languages, and to changes in languages over time. Given the compelling nature of hierarchical descriptions, it is tempting to conclude that the process of speech production involves using hierarchical relations among linguistic units to specify the appropriate order of the units to be produced. However, speech production requires the participation of a variety of motoric and other information-processing systems (Bock, 1982) that might influence the manner in which order is represented. In addition, it involves a number of phases, including the planning and the execution of an utterance. While the hierarchical relations among linguistic units undoubtedly contribute to the early planning phases, the control of units in the latter phases is not fully understood.

We will discuss the ordering of speech units in terms of utterance plans, representations specifying the order of the segments to be produced. It is generally, though not universally (Kelso, Tuller, & Harris, 1983), accepted that movement of the articulators is guided by an utterance plan or motor program, which represents the utterance to be produced. Formulations of this hypothesis have been put forth as "action schemas" (Lashley, 1951), "motor plans" (Miller, Galanter, & Pribram, 1960), and "motor programs" (Keele, 1968; Sternberg, Monsell, Knoll, & Wright, 1978). The plans or programs are thought to contain at least some of the information necessary to specify the utterance (e.g., ordering of segments, articulatory specification of segments, relative timing of segments, etc.). We have chosen the term "utterance plan" over the more common and controversial term "motor program" because there is no clear-cut basis for deciding whether a given type of information ought to be considered "motoric."

An utterance plan can be viewed as a representation mediating between two phases of speech production: planning and execution (cf. Rosenbaum, Kenny, & Derr, 1983). The planning phase involves con-struction of utterance plans. Planning processes can occur before overt movement begins, and do not necessarily result in any movement at all. Execution processes act on the utterance plan to produce action. Thus, they are necessarily linked to overt movement. In this paper, we will attempt to determine how information about the serial order of syllables is represented in utterance plans.

MODELS OF SERIAL ORDER

We will consider three ways in which serial-order information might be represented in utterance plans: element-to-element coding, element-to-position coding, and hierarchical coding. The first two types of representations are linear structures in which lower level elements are not associated with any higher level elements. In contrast, a hierarchical program represents elements at different levels of abstraction. The order of the lower level units is specified by their relations to higher level units, not by their relation to each other or by their absolute position in the utterance. The models to be reviewed here are based on various types of data, and often address different levels of speech production. They have been grouped according to the different types of serial-order coding, either because they rely heavily on one type of coding, or because they provide particularly good illustrations of a type of coding. In actuality, most of the models are hybrids, making use of more than one kind of serial-order coding. Through this review, we hope to show that each type of serial-order coding may be viable as an account of some aspects of the ordering of linguistic units. The focus will be on how order is represented, not on why linguistic units appear in a given order.

Element-to-Element Coding Models

According to element-to-element coding models, the serial order of linguistic units is represented through direct associations between adjacent units in the utterance to be

produced. These models have their roots in associative chain theory, and as such, have very few proponents in contemporary cognitive psychology. Lashley (1951) argued that this kind of model is unable to account for a variety of aspects of language production, such as relations between noncontiguous linguistic units, naturally occurring speech errors, and the rapidity of articulation.

Wickelgren (1969) proposed the only prominent element-to-element model. It uses "context-sensitive allophones" to represent the order of phonemes in words. For every phoneme, the model assumes that an allophonic representation exists for each context in which it can appear. Thus, the /b/'s in "able" and "oboe" would be represented differently; the first as $/_{e}b_{l}/$ and the second as $/_{o}b_{o}/$. All phonemes in all neighboring contexts are assumed to be stored in associative memory. Wickelgren's model belongs to the element-to-element class because each sound in a sequence is initiated through activation received from the preceding segment.

Wickelgren justified his model largely on grounds of sufficiency. It does specify a mechanism that could produce serially ordered sounds in words. However, Wickelgren's model, in its original form, failed to predict a host of regularities in speech errors, such as the tendency of segments to substitute for phonetically similar segments, and the tendency of segments to move into an identical syllable position (MacKay, 1970). (A revised model, Wickelgren (1976), can account for the effects of syllabic position.) While Wickelgren's model does not seem to be viable as a sole account of phoneme order in words, the distinction between planning and execution processes raises the possibility that different ordering schemes are used during different phases of production. Errors in performance could occur during a planning phase that uses hierarchical linguistic relations (Rosenbaum et al., 1983), while later execution phases might still use element-to-element coding. If this were the case, then a possible role for a model like Wickelgren's would remain.

Element-to-Position Coding Models

In element-to-position (henceforth "position") coding models, a speech segment maintains its appropriate place in a sequence through an association to a member of a naturally ordered set (cf. Conrad, 1965). The ordering in the set then provides a way of sequencing the segments properly. Models of this sort have been proposed as accounts of speech-latency data (Sternberg et al. 1978) and speech-error patterns (Shattuck-Hufnagel, 1979).

Sternberg et al. (1978) proposed a model that describes the output of word-size units during a phrase-length utterance. Their model accounts for results from an experimental procedure in which subjects produced an utterance as quickly as possible in response to a signal. They found that the latency to initiate the utterance increased linearly with its length measured as the number of stressed syllables in the utterance. Furthermore, Sternberg et al. (1978) found that the duration of the utterance was a quadratic function of the number of stressed syllables, with the quadratic component being equivalent to the slope of the latency function.

Their account of these data deals, in part, with how the stressed syllables are output in the proper order. According to their model, outputting a stressed syllable involves searching a buffer with subprograms corresponding to the stressed syllables. Search time is directly related to the number of subprograms in the buffer. One result obtained by Sternberg et al. (1978) indicates that this search involves position tags. They found a linear effect of number of stressed syllables on latency even when all of the syllables were identical, implying that each search is for a subprogram in a specific position.

Shattuck-Hufnagel (1979) presented a model similar to that of Sternberg et al.

(1978) as an account of naturally occurring speech errors. The model describes how the ordering of words and phonemes is represented for an utterance. It assumes that a set of ordered word slots is generated and a set of lexical items is retrieved from long-term memory and placed in a separate short-term buffer. A "scan-copier" scans the buffer for the first word and then copies it into its slot. A checkoff monitor marks the selected target. The process then proceeds in the same manner for the second word. After the words have been copied into the ordered slots, they can be output. An analogous set of processes operates at the phoneme level. Shattuck-Hufnagel's model is a position-coding model because the final ordering of the segments is represented by their placement in slots corresponding to serial positions in the utterance.

The above position models offer a very simple way of representing order information. Position-coding schemes do not seem in principle incompatible with any known characteristics of speech production. However, they may need substantial elaboration to account for a wide variety of phenomena. This might take the form of blending hierarchical coding with position coding to achieve a hybrid model.

Hierarchical Coding Models

Hierarchical coding models constitute the largest and richest variety of speech-production models. There are several reasons for this. As we pointed out previously, linguistic analysis provides a hierarchical description for the units of language, which lends itself naturally to representing serial order in speech production. Similarly, cognitive psychology has made wide use of hierarchical models in attempting to account for a variety of other mental functions. Here, we will review several of these hierarchical models. Some deal directly with speech production, while others are concerned with sequence production more generally.

General sequence production. Greeno and Simon (1974) performed a formal analysis of how sequential behavior can be produced from a stored hierarchical code. Their analysis assumed that this code is the input to a generative or interpretive process. Hierarchical codes were defined as ones where any acceptable transformation could be applied to a single element or to a sequence of elements. Greeno and Simon (1974) considered three transformations or operators: repetition (R), mirroring (M), and transposition (T). The repetition operator causes a certain element or sequence of elements to be repeated, the mirroring operator causes a sequence of elements to be repeated in reverse order, and the transposition operator causes a given element or a sequence of elements to be repeated after being transformed by some increment. The embedded structure in this kind of hierarchical representation allows a very long sequence to be represented very succinctly.

Collard and Povel (1982) and Rosenbaum et al. (1983) have extended the Greeno and Simon style of modeling to generate predictions about inter-response times in sequence production. They postulated that a sequence is ordered by a hierarchical tree structure, with the levels in the hierarchy specified by the transformations described in Greeno and Simon (1974). Producing a sequence involves traversing the tree until a terminal node (i.e., an element to be produced) is encountered. In a task involving sequential keypresses with different fingers, Rosenbaum et al. (1983) found that inter-response intervals were a linear function of the distance traveled in the postulated tree. The hierarchical structure consisted of a tree with binary branches induced by including obvious patterns of repetition and alternation among the fingers.

The work of Greeno and Simon (1974), Collard and Povel (1982), and Rosenbaum et al. (1983) is not directly relevant to speech production. The formal analyses (Greeno & Simon, 1974) derived their

power from restrictions on the kinds of structures to which they are applied. These analyses are only applicable when transformations can be used at any level of the hierarchy. Natural speech sequences, described at the level of words or phonemes, do not typically have structure of this exact sort and thus would not in general use ordering based on nominal transformations such as repetition and alternation.

On the other hand, this body of work does show the organizational power of hierarchical coding and provides an empirical demonstration of its use under certain circumstances. Moreover, there is evidence that speech production is sensitive to the kinds of transformations described above. Schourup (1973) has shown that some tongue twisters are best described in terms of conflicting patterns of repetition and alternation. In addition, factors other than the identity of the elements might provide a basis for grouping speech segments into hierarchical structures. Specifically, stress and timing have been described as deriving from a hierarchical tree with binary branches (Liberman & Prince, 1977; Martin, 1982). Hierarchical coding of speech sequences, based on principles of grouping in terms of stress, might have properties similar to those found for keypresses.

A second kind of hierarchical ordering, which differs from the operator conception of Greeno and Simon (1974), involves the notion of "control elements" (Estes, 1972). A control element is a higher level representation that specifies and controls the ordering of lower level elements. As such, it solves the order problem at one level by creating additional representations at higher levels. Estes' (1972) model attempts to describe short-term memory for order by asserting that the elements to be remembered are refreshed in sequence through a superordinate control element. This refreshing is designed to preserve the relative temporal positions of the items. Perturbations in the cycling between elements are

responsible for the loss of order information. Estes (1972) further argues that his model can be naturally extended to account for long-term memory of order information and for sequence production, by establishing inhibitory links between elements to prevent the perturbation in cycling that causes forgetting. The use of inhibition between elements makes this a hybrid model, using both hierarchical and element-to-element coding.

Hierarchical models of speech production. Several hierarchical models involving control elements have been proposed for speech production. MacKay (1982) proposed the most encompassing of these, arguing that hierarchical models can explain both the fluency and flexibility of skilled performance. His system begins with nodes (control elements) at the level of propositional concepts and terminates with muscle-movement nodes. Many levels of nodes intervene between these levels, each controlling the ordering of activation in the next lower level. The intermediate levels include conceptual compound nodes, lexical concept nodes, syllable nodes, phonological compound nodes, phonological nodes, and feature nodes. The major sources of evidence presented by MacKay (1982) to support this system were practice effects on speech rate and transfer of practice effects to novel utterances. Improved performance with practice may be attributed to the strengthening of connections between nodes at different levels in the proposed hierarchy. Studying the transfer of practice offers a potential way of isolating the level of representation at which a given practice effect occurs. The basic appeal of MacKay's (1982) model stems from its attempt to provide an overarching theory of speech production. As a negative consequence, the model lacks detailed justification for some of the levels proposed. Using the transfer data, MacKay (1982) was able to justify some of his levels, but not others.

Dell and Reich (Dell, 1984, 1986; Dell &

Reich 1980, 1981) have also applied the notion of hierarchical coding to speech production. Like MacKay (1982), they have addressed a number of linguistic levels including semantic and syntactic, but have focused most fully on the lexical, syllabic, and phonemic levels. A simulation of their model (Dell & Reich, 1980; Dell, 1986) included representations at the word, syllable, and phoneme levels. Representations at each level consisted of nodes connected in a hierarchical arrangement. Each node transmitted excitatory activations to neighbors at lower and higher levels. The model produced speech errors when activations for a given node exceeded threshold at an inappropriate time. Comparison of these errors to human speech errors, both naturally occurring and artificially elicited, have provided the major empirical tests of the model. It successfully accounted for the major classes of human speech errors, as well as several important subtleties.

Evaluation of hierarchical models. As noted above, the class of hierarchical models is extremely diverse, and thus a general assessment of their advantages and disadvantages is difficult to give. The formal characterization of sequence production by Greeno and Simon (1974) offers a way to specify what levels would be present in hierarchical trees such as those postulated by Collard and Povel (1982) and Rosenbaum et al. (1983). However, the levels present in a linguistic hierarchy are not determined by factors such as repetition of lower level elements, but rather by semiarbitrary groupings of lower level elements (e.g., phonemes) into higher level elements (e.g., words).

The control-element models of MacKay (1982) and Dell and Reich (Dell, 1984; Dell & Reich, 1980, 1981) incorporate linguistically specified units. The direct incorporation of hierarchical linguistic relations into a production model has some appeal, since it is likely that these relations must play some role in speech production. Nevertheless, there are at least two reasons for doubting that this type of model constitutes a complete solution to the problem of serial order in speech production. First, whether or not they can actually be made to work is by no means certain. Dell and Reich (1981) have successfully simulated certain aspects of their model, but it is not clear that a more ambitious model such as MacKay's (1982) will actually work as currently described. A second criticism of these models is that they make no distinction between planning and execution processes. They do allow for advance preparation of a response in terms of subthreshold activation. However, they do not postulate a short-term utterance plan: a construct that may be required to account for certain empirically established speech-production phenomena (Sternberg et al., 1978).

RESEARCH RATIONALE

The goal of the present research is to determine which of the three types of models outlined above best describes how syllable order is represented in fully prepared utterance plans. There are arguments in favor of all three models, and no data that definitively favor one over the others. The results of linguistic analysis and phenomena such as speech errors indicate that hierarchical factors play an important role in the production of speech. However, it is not clear that this role extends beyond an early planning phase to shape the representations used in a fully prepared utterance plan or motor program (Rosenbaum et al., 1983; Sternberg et al., 1978).

At the level of fully prepared utterance plans, simple linear structures, as embodied in element-to-element coding and element-to-position coding, potentially have at least two significant advantages over hierarchical structures. A primary function of the utterance plan could be to represent information for the motor system in an easily accessible fashion (Sternberg et al., 1978). By including only a single level of representation, element-to-element

coding and element-to-position coding would make information more readily accessible than hierarchical coding does, thereby allowing greater speed and accuracy of output. In addition, element-to-element and element-to-position coding might allow two or more speech segments to be articulated at a time. Anticipatory coarticulation has been found to occur over as many as seven phonemes, and takes place across syllable and word boundaries (Kent & Minifie, 1977). The most successful accounts of coarticulation (and of its analog in typing) use a process of look-ahead through a linear sequence of intended targets (Henke, 1966; Rumelhart & Norman, 1982). This provides additional motivation for testing element-to-element, and element-to-position coding models as serial-order representations in fully prepared utterance plans.

In order to test the three types of models, we must develop concrete instantiations of each. Both element-to-element coding and position coding can be specified in a straightforward way. However, the rich variety of hierarchical models creates a more difficult task. In the following experiments, we examine a hierarchical model based on a tree with binary branches, as shown in Figure 1. This model is very similar to the one developed by Rosenbaum et al. (1983) for manual performance. Ac-
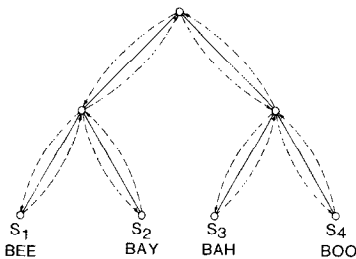

FIG. 1. A hierarchical structure specifying the order of a sequence of syllables. The terminal nodes specify the identities of the syllables to be produced. The order of these nodes is specified by the higher order nodes to which they are connected. The dashed lines illustrate how a tree-traversal mechanism (Rosenbaum et al., 1983) might operate on the hierarchy in order to output the syllables.

cording to the model, the syllables to be produced (e.g., BEE–BAY–BAH–BOO) are connected to terminal nodes at the bottom level of the hierarchy. The terminal nodes are not connected directly to each other. Instead, their order is represented indirectly by their connections to the higher level nodes. The production of correctly ordered speech could be accomplished by a procedure such as traversing the tree structure from one terminal node to the next, as indicated by the dashed lines in Figure 1 (cf. Rosenbaum et al., 1983). Other output procedures besides tree-traversal are also possible. For example, activation could spread from the higher nodes to the lower ones with greater amounts going down the left-hand branches. When the unit corresponding to a terminal node reached threshold, it would be output, and its activation released to subsequent units (cf. Dell & Reich, 1982; Rumelhart & Norman, 1982, for similar proposals).

Previous theoretical and empirical work supports the notion that binary tree structures may play a role in manual responses and in speech production. That work motivates the present focus. As described above, Rosenbaum et al. (1983) studied the nature of response plans underlying production of a sequence of keypresses described by a binary tree. In addition, work on stress and the rhythmic organization of speech provides a basis for investigating the role of binary tree structures (Liberman & Prince, 1977; Martin, 1972). It has been argued that stress and rhythm are not properties of individual linguistic units; rather they depend on relations among the units of an utterance, and they stem from a binary tree representation of the syllables in an utterance. Such accounts of stress and rhythm (Liberman & Prince, 1977; Martin 1972) imply a potential role for binary tree structures in the control of prosody. This raises the possibility that the same binary tree representations might underlie the control of serial order.

Given our concern with prosodically mo-

tivated binary tree structures, it is reasonable to focus on the ordering of syllables, since they are the units of speech believed to carry stress (Selkirk, 1984). Furthermore, syllables constitute a very natural level for analyzing serial order because they are the smallest units of speech that can be pronounced by themselves (Stetson, 1951). Successive syllables can logically be independent of one another, making it possible to examine the nature of ordering mechanisms that are imposed as part of production and that might be separate from preexisting linguistic groupings, such as those specified by words.

Response-priming procedure. The present experiments use a response-priming procedure (Meyer and Gordon, 1985; Meyer, Sternberg, Knoll, & Wright, 1978; Rosenbaum & Kornblum, 1982) to distinguish between the three types of serial-order models. This procedure involves a choice reaction-time task in which a subject must rapidly make one of two vocal responses (a "primary" utterance or a "secondary" utterance) depending on which of two response signals is presented. Unequal emphasis is placed on the production of the two utterances. The subject has to say the primary utterance as quickly as possible in response to the primary response signal (a high-pitched tone). Consequently, it is to the subject's advantage to prepare fully for the primary utterance at the start of each trial. However, if the secondary response signal is presented, the subject must switch and produce the secondary response.

The preparation of the primary response presumably involves constructing an utterance plan representing many aspects of the particular utterance to be produced. If the subject must instead produce the secondary response, this plan may be edited so that it is appropriate for the newly required utterance (Rosenbaum, Inhoff, & Gordon, 1984). The amount of necessary editing will be reflected by performance on the secondary response.

The different models of serial-order representation describe different pairs of primary and secondary responses as having greater or lesser overlap in terms of how their constituents are ordered. Assuming that the editing process exploits this relation, performance on a secondary response should improve as its overlap with the primary response increases. The latency to initiate the secondary response provides one measure of performance that reveals how much editing or reprogramming is required to switch from the primary to the secondary response.

The duration and accuracy of the secondary response may also provide important evidence about the representation of serial-order information. A subject may not complete all the necessary reprogramming before beginning a secondary response. If this happens, then reprogramming might continue during the utterance, leading to increased response duration. Moreover, critical aspects of programming may not be done at all if the subject rushes the response too much, resulting in diminished response accuracy. Thus, response duration and accuracy complement response latency as dependent measures that may reveal how the production of the secondary response depends on overlap between the serial-order representations of the primary and secondary responses.

EXPERIMENT 1

Experiment 1 used the response-priming procedure to examine the hierarchical binary-tree model discussed earlier. It explored whether speaker's spontaneously adopt such a representation when producing a four-syllable utterance. It also explored whether the presence of some structure in the utterances influences the manner of coding. To construct the syllable sequences, we manipulated the relationship between the primary and secondary syllable sequences. In one condition, the ordering of the secondary syllable sequence was "hierarchically congruent" with the primary syllable sequence. This was done

by having a secondary syllable sequence that preserved the lower branches of the tree structure controlling the primary syllable sequence, as shown in Figure 2. For example, if the primary syllable sequence was $S_1 S_2 S_3 S_4$ (e.g., BEE–BAY–BAH–BOO), then a hierarchically congruent secondary sequence would be $S_3 S_4 S_1 S_2$ (e.g., BAH–BOO–BEE–BAY). In another condition, the secondary and primary syllable sequences were hierarchically incongruent. For example, the sequence $S_3 S_2 S_4 S_1$ (e.g., BAH–BAY–BOO–BEE) would comprise a hierarchically incongruent sequence.

Referring to Figure 2, it can be seen that the control structure for the hierarchically congruent secondary sequences were derived from the structure for the primary sequence by simply reversing the direction of the branches from the top node. In contrast, deriving a control structure for the incongruent sequences would require a great deal more rearranging of the primary control structure. Thus, if syllable order is controlled by a binary tree structure, we would expect that performance on the hier-

archically congruent secondary syllable sequences would exceed performance on hierarchically incongruent secondary syllable sequences.

A second factor manipulated in Experiment 1 was the pattern of initial consonants used in the syllable sequences. In one condition, all of the syllables of a sequence began with the same consonant (e.g., BEE–BAY–BAH–BOO). We refer to these as *homogeneous syllable sequences*. In the other condition, the initial consonants of the primary sequences alternated between two different consonants (e.g., BEE–DAY–BAH–DOO). We refer to these as *heterogeneous syllable sequences*. By using these different types of sequences, we hoped to determine whether subjects' adoption of a hierarchical control structure necessarily depends on a pattern of alternation in the syllable sequences to be produced.

Method

Subjects. Seven University of Michigan students served as paid subjects. They were randomly chosen from the Human Performance Center subject pool and were paid $4.00 an hour plus a bonus based on the quality of their performance.

Apparatus. A DEC PDP-11/34 computer controlled the experiment. Visual messages, syllable sequences, and response feedback were presented to the subject on a video terminal (HP 2621A). Warning tones and response tones were presented over headphones (Sennheisser HMD 224). A microphone attached to the headphones transmitted subjects' vocal responses to an A-to-D converter, which sampled them at a 5 KHz rate. A software voicekey (Gordon & Meyer, 1984; Meyer & Gordon, 1985) was used to measure the latency and duration of each utterance. The experimenter, blind to the correct response, typed the subject's response into the computer. This allowed the computer to evaluate the accuracy of the vocal responses.
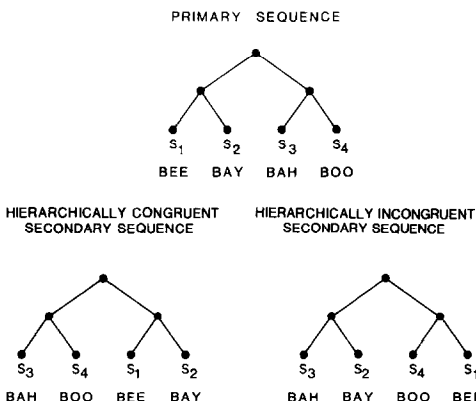
Design. Each subject was tested for five

FIG. 2. Hierarchical coding structures for a primary syllable sequence, a hierarchically congruent secondary syllable sequence, and a hierarchically incongruent secondary sequence. The "S"'s indicate syllables, and their subscripts indicate the syllables' position in the primary sequence. The coding structure for the hierarchically congruent secondary sequence can be obtained from the coding structure for the primary sequence by simply reversing the branches from the top node. Obtaining the coding structure for the hierarchically incongruent secondary sequence would require many more changes.

sessions on separate days within a 10-day span. There were 8 trial blocks per session, each of which included 24 trials. Half the trials required the primary response, and half required the secondary response. Table 1 shows the syllable sequences assigned to the different response types. On a given trial block, one primary syllable sequence and its accompanying secondary sequences were used. Across the eight blocks of a session, each of the sets of primary and secondary syllable sequences was used. The order of block types was counterbalanced across subjects and sessions.

Procedure. At the beginning of each trial, the subject was shown a primary-response sequence and a secondary-response sequence. They appeared as two rows of syllables with the primary response on top of the secondary response. The syllable sequences remained on the screen for 2 s during which the subject was supposed to prepare to produce the primary response. The primary syllable sequence then disappeared while the secondary sequence remained visible. After a 1-s pause, the subject heard three medium-pitched (417 Hz) tones of 100 msec duration, separated by 400 msec of silence. A response tone was presented 400 msec after the last warning tone. On half the trials, the response tone was high pitched (833 Hz), indicating that the subject should produce the primary syllable sequence. On the other half of the

trials, the response tone was low pitched (208 Hz), indicating that the secondary sequence should be produced. The subject did not know beforehand which response tone would actually occur, but was supposed to prepare for the primary one in any event.

The subject received negative feedback immediately after a trial if she or he produced an incorrect response. In addition, the message "RESPONSE TOO SLOW" was given on trials requiring a primary response if the subject's completion time (the latency to initiate the utterance plus the utterance's duration) exceeded a deadline. The deadline was initially set at 1200 msec, and was lowered as the subject's performance improved. Subjects' final deadlines fell in the range of 750 to 900 msec. At the end of a trial block, the subject was shown mean completion times for primary and secondary responses and number of errors. A point total, calculated with a formula described in Meyer and Gordon (1985), was also shown to the subject. A running tally of the block scores was kept so that the subject could track his or her performance.

Results

The principal results are shown in Table 2, which displays mean latencies, durations, and error rates for primary and secondary responses in the various conditions. Separate analyses of variance were per-

TABLE 1
SYLLABLE SEQUENCES USED IN EXPERIMENT 1

	Secondary sequences	
Primary sequences	Hierarchically congruent	Incongruent
1234	3412	3241
1536	3615	3561
2143	4321	4132
5163	6351	6135
3214	1432	1243
3516	1635	1563
4123	2341	2134
6153	5361	5136

Note. 1 = BEE(/bi/); 2 = BAY (/be/); 3 = BAH (/ba/); 4 = BOO (/bu/); 5 = DAY (/de/); 6 = DOO (/du/).

formed on these measures to test for differences between primary and secondary responses. Subjects produced considerably shorter latencies on primary responses (mean latency = 297 msec) compared to secondary responses (mean latency = 588); $F(1,6) = 40.3$, $p < .01$. Mean durations were also shorter for primary responses (454 msec) than for secondary responses (516 msec); $F(1,6) = 16.4$, $p < .01$. Finally, error rates were lower for primary responses (3.6%) compared to secondary responses (7.5%); $F(1,6) = 8.3$, $p < .05$. Thus, all three measures showed significantly better performance on primary responses than secondary responses.

The effects of primary–secondary relationship on secondary-response performance were assessed through separate analyses of variance on response latency, duration, and error rate. Latencies were shorter for hierarchically congruent secondary responses (576 msec) than incongruent secondary responses (600 msec); $t(6) = 3.2$, $p < .02$. The mean duration was less for hierarchically congruent secondary responses (501 msec) than for incongruent secondary responses (530 msec); $t(6) = 5.2$, $p < .01$. And finally, error rates were lower for hierarchically congruent secondary responses (4.9%) than for incongruent secondary responses (10.1%); $t(6) = 2.5$, $p < .05$.

We also tested for differences in performance between homogeneous versus heterogeneous responses. The only significant effect was that homogeneous responses had longer durations than heterogeneous responses; $t(6) = 6.7$, $p < .01$. Analysis of error rates showed a marginally significant effect, with fewer errors occurring for the homogeneous strings; $t(6) = 2.1$, $p < .10$. More importantly, there were no significant interactions between heterogeneity and hierarchical congruence for any of the three dependent measures ($p > .10$ in all cases).

Analysis of primary-response performance also showed significant main effects of primary/secondary relationship and of sequence homogeneity. Fewer errors occurred for hierarchically congruent primary responses than for incongruent responses; $t(6) = 3.8$, $p < .01$. Latencies were also marginally shorter for congruent primary responses than incongruent primary responses; $t(6) = 2.0$, $p < .10$. Homogeneous responses had shorter latencies than heterogeneous responses; $t(6) = 4.0$, $p < .01$. However, homogeneous responses had longer durations than heterogeneous responses; $t(6) = 12.1$, $p < .01$.

In addition to analyzing error rates, we also examined the types of errors that subjects made. Of 136 secondary-response errors, 63 were "pure" order errors; the correct syllables were produced, but in the

TABLE 2
RESULTS OF EXPERIMENT 1

	Primary/secondary relationship			
	Homogeneous		Heterogeneous	
	Hier. congr.	Incongr.	Hier. congr.	Incongr.
Latency (ms)				
Secondary	588	608	564	593
Primary	287	290	301	309
Duration (ms)				
Secondary	537	559	465	501
Primary	504	508	403	399
Error rate (%)				
Secondary	3.0	9.9	6.7	10.3
Primary	2.8	4.3	2.6	4.6

wrong order. The remaining 73 errors contained an assortment of responses with omitted syllables, repeated syllables, and garbled speech. Of the pure order errors, 38 involved an interaction between two syllables. Of these, 65.8% involved exchanges between syllable positions three and four. This is important because in short-term memory experiments with four elements, the most frequent error type is an exchange of positions two and three (Estes, 1972), not positions three and four. Exchanges between positions two and three accounted for only 7.9% of the order errors we observed. Given the limited number of order errors, it was not possible to determine how primary/secondary relationship might have influenced error types.

Discussion

Evidence of primary-response preparation. The principal prerequisite for drawing inferences from the response-priming procedure is that subjects must achieve a high state of preparation for producing the primary responses. All three measures of performance (latency, duration, and error rate) were clearly superior for primary responses compared to secondary responses. This indicates that subjects heeded our instructions and differentially prepared to produce the primary responses.

The absolute levels of performance on primary responses are also helpful in assessing subjects' level of preparedness. If subjects were highly prepared to produce a primary response, then we would expect their performance to be similar to that obtained in a simple reaction-time task, where no uncertainty exists about the required response. The primary responses in Experiment 1 were quite similar to ones used by Sternberg et al. (1978, p. 128) in one condition of a simple reaction-time task. They obtained a mean latency of approximately 292 msec compared to our mean latency of 297 msec. Their mean duration was approximately 515 msec as compared to 454

msec in the present experiment. Although the two sets of results can not be strictly compared, because they involved different subjects and materials, it is clear that performance on our primary responses was in the same range as in a simple reaction-time task. This indicates that subjects probably programmed the primary response before the response signal.

Evidence for hierarchical coding. All three measures (latency, duration, and error rates) showed better performance for hierarchically congruent secondary syllable sequences than for incongruent sequences. In addition, fewer errors and somewhat shorter latencies were found for primary responses paired with hierarchically congruent secondary responses than for those paired with incongruent secondary responses. Modifying an utterance is apparently easier if the resulting plan preserves aspects of the structure of the original utterance plan. This supports the notion that producing a prepared four-syllable sequence is guided by a tree structure as illustrated in Figure 1.

We found no evidence of an interaction between hierarchical congruence and type of syllable sequence (heterogeneous versus homogeneous). It appears that the adoption of a hierarchical structure for controlling serial order is not dependent on the presence of salient subgroupings in a sequence. This finding is significant, since previous demonstrations of hierarchical coding have required such subgroupings to induce effects (Rosenbaum et al., 1983). The results of Experiment 1 suggest that hierarchical coding of order may be spontaneously adopted in articulation of syllable sequences, even if no obvious structure is present to bias those sequences.

EXPERIMENT 2

While the results of Experiment 1 may be interpreted as supporting the hierarchical coding model, an alternative explanation is also plausible. The syllable se-

quences that we characterized as hierarchically congruent were likewise congruent in terms of the order relations prescribed by an element-to-element coding model. In particular, the hierarchically congruent secondary sequences (i.e., $S_3S_4S_1S_2$) preserved two of the element-to-element associations (i.e., S_3–S_4 and S_1–S_2) from the primary syllable sequences. Thus, the facilitation observed in Experiment 1 could be due to element-to-element coding rather than hierarchical coding of order information.

The goal of Experiment 2 was to test between the hierarchical and element-to-element coding models. In order to do this, we introduced a new kind of secondary syllable sequence that had element-to-element congruence with the primary syllable sequence, but did not have hierarchical congruence. The order of syllables for the new secondary sequences was $S_2S_3S_4S_1$ (e.g., a primary response of BEE–BAY–BAH–BOO changed to a secondary response of BAY–BAH–BOO–BEE). As can be seen in Figure 3, these sequences maintained two of the element-to-element associations present in the primary sequence. Thus, they should yield relatively good performance if element-to-element coding is used
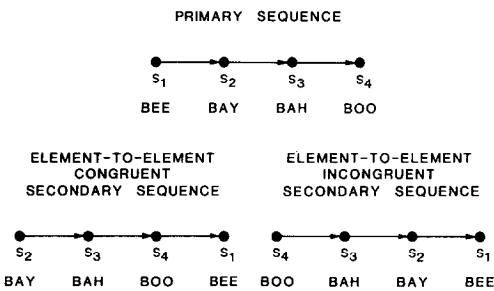
PRIMARY SEQUENCE

FIG. 3. Element-to-element coding structures for a primary syllable sequence, a congruent secondary sequence, and an incongruent secondary sequence. The "S"s indicate syllables and the arrows indicate directional links. The coding structure for the congruent secondary sequence can be obtained by simply detaching the first syllable of the primary sequence and reattaching it to the end. Obtaining the coding structure for the incongruent sequence would require reversing the directions of all the directional links.

to represent serial order. However, they would cause substantial disruption of a binary-tree structure for producing the primary response (cf. Figures 1 and 2). We also included hierarchically congruent secondary responses, such as those used in Experiment 1 ($S_3S_4S_1S_2$), and incongruent secondary responses represented by the sequence $S_4S_3S_2S_1$. The experiment only used homogeneous syllable sequences, whose members all began with the same consonant (i.e., /b/).

Method

Subjects, apparatus, and procedure. Eleven students, none of whom had participated in the previous experiment, served as paid subjects. The apparatus and general procedure were the same as in the previous experiment.

Design. The syllable sequences are shown in Table 3. Assignment of the secondary sequences to the various primary/secondary order relations was counterbalanced across primary-response sequences, thereby controlling for intrinsic differences in the difficulty of producing the secondary responses. As in Experiment 1, a single primary sequence, and its accompanying secondary sequences, were used in a given block of trials. The primary and secondary sequences varied across blocks. There were 8 blocks per session, each containing 12 trials. The order of blocks varied across subjects and sessions. Subjects participated in one practice session followed by four test sessions. The data from practice sessions were not analyzed.

Results

Table 4 shows the mean latencies, durations, and error rates for primary and secondary responses with the various serial-order relations. As before, latencies were significantly shorter for primary responses (296 msec) than for secondary responses (652 msec); $F(1,10) = 86.9, p < .01$. Durations were also shorter for primary re-

TABLE 3
SYLLABLE SEQUENCES USED IN EXPERIMENT 2

	Secondary sequences		
Primary sequences	Hier. congr.	Element-to-element congr.	Incongr.
1234	3412	2341	4321
4123	2341	1234	3214
1432	3214	4321	2341
2341	4123	3412	1432
2143	4321	1432	3412
3412	1234	4123	2143
4321	2143	3214	1234
3214	1432	2143	4123

Note. 1 = BEE (/bi/); 2 = BAY (/be/); 3 = BAH (/ba/); 4 = BOO (/bu/).

sponses (521 msec) than for secondary responses (615 msec); $F(1,10) = 19.0$, $p < .01$. Similarly, error rates were lower for primary responses (4.3%) than for secondary responses (11.1%); $F(1,10) = 14.8$, $p < .01$.

Planned comparisons were performed on secondary response performance as a function of the primary/secondary order relation. A significant difference was found between the error rates for hierarchically congruent responses (7.8%) and incongruent responses (14.6%); $t(20) = 2.5$, $p < .02$. Consistent with this finding, the latencies for the hierarchically congruent responses were marginally shorter than those for the incongruent responses (641 versus 662 msec; $t(20) = 2.0$, $p < .10$). On all

three measurements, the hierarchically congruent responses also showed better performance than did responses with element-to-element congruence. When the significance levels associated with these contrasts were combined using Fischer's combination of tests, the difference between responses with hierarchical and element-to-element congruence came very close to reaching standard levels of significance; $\chi^2 (6) = 12.1$, $p = .06$. No differences were found in latency, duration, or error rates for performance on secondary responses with element-to-element congruence compared to the incongruent secondary responses ($p > .10$ in all cases).

Contrasts involving primary responses were also evaluated in order to test for pos-

TABLE 4
RESULTS OF EXPERIMENT 2

	Primary/secondary relationship		
	Hier. congr.	Element-to-element congr.	Incongr.
Latency (ms)			
Secondary	641	652	662
Primary	298	297	292
Duraton (ms)			
Secondary	607	618	619
Primary	520	524	519
Error rate (%)			
Secondary	7.8	10.9	14.6
Primary	4.9	3.7	4.2

sible effects of the primary/secondary order relations. No significant differences were found for any of the comparisons based on any of the dependent measures; $p > .10$ in all cases.

A detailed analysis of errors was performed as in Experiment 1. Of 175 errors, 65.7% were order errors; the correct syllables were produced in an incorrect order. The remaining errors involved syllable omissions, repetitions, or aborted responses. The proportion of order errors did not vary significantly as a function of the primary/secondary order relation; $\chi^2 (2) = 2.0, p > .10$. The order errors were further broken down according to which syllables changed places. As found earlier, the most frequent exchange involved positions three and four, accounting for 34.8% of the order errors. Two–three exchanges were the next most frequent, accounting for 14.8% of the errors. No other particular kind of error occurred with sufficient frequency to merit mention.

Discussion

The results of Experiment 2 provide additional support for a hierarchical coding model of serial order at the level of fully prepared utterance plans. We found significant facilitation for the production of secondary syllable sequences with hierarchical congruence compared to incongruent syllable sequences. Less facilitation was observed for the production of secondary syllable sequences with element-to-element congruence. This indicates that the facilitation for the hierarchically congruent sequences was not due simply to the presence of two element-to-element associations from the primary syllable sequence. If the element-to-element associations were responsible for the facilitation, as much facilitation should have occurred for the secondary sequences with element-to-element congruence, because those sequences also maintained two element-to-element associations from the primary syllable sequence.

Because performance levels on secondary responses with element-to-element congruence were intermediate between hierarchically congruent and incongruent responses, we can not rule out element-to-element coding as contributing to the representation of serial order. However, such a coding scheme is insufficient as a sole account of the results of Experiment 2, suggesting a role for the binary tree structure in the representation of serial order.

EXPERIMENT 3

Experiment 3 tested whether element-to-position coding contributes to serial-order representation. Element-to-position coding represents order by association between items (e.g., syllables) to be produced and members of an ordered set, analogous to placing items in prearranged slots, like those of a mailbox (Conrad, 1965). To check whether this type of representation plays a role in the production of fully prepared utterances, we used secondary syllable sequences that preserved two element-to-position associations from the primary sequence. For example, the sequence $S_3S_2S_1S_4$ maintained the serial positions of two syllables, and as such had element-to-position congruence. If the production process involves some direct element-to-position coding as illustrated in Figure 4, then one might expect such sequences to yield some facilitation of performance. In addition, we included hierarchically congruent sequences, like those from the previous two experiments, and we included incongruent sequences.

Method

Subjects, apparatus, and procedure. A new sample of 10 subjects participated in this experiment. The apparatus and procedure were the same as before.

Design. Table 5 shows the chosen syllable sequences. As before, a single primary sequence, and its accompanying secondary sequences were used in a given
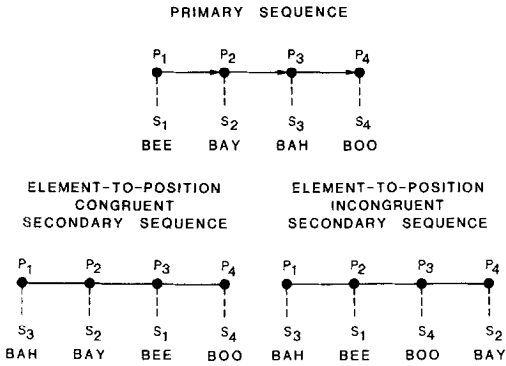
PRIMARY SEQUENCE

FIG. 4. Element-to-position coding structures for a primary syllable sequence, a congruent secondary syllable sequence, and an incongruent secondary sequence. The "P"'s indicate positions, and the "S"'s indicate syllables that are linked to the positions. The coding structure for the congruent secondary sequence can be obtained by changing the positions of syllables 1 and 3. Obtaining the coding structure for the incongruent sequence would require changing the positions of all syllables.

block of trials. Six blocks of trials, one with each set of sequences, comprised a session. The order of blocks varied across subjects and sessions. Each block contained 24 trials. Subjects participated in five sessions. The first session served as practice and was excluded from the data analyses.

Results

The mean latencies, durations, and error rates for primary and secondary responses with the various serial-order relations are

shown in Table 6. Performance on the primary responses was significantly better than on the secondary responses; latencies were shorter ($F(1,9) = 70.9, p < .01$), durations were shorter ($F(1,9) = 17.4, p < .01$), and error rates were lower ($F(1,9) = 24.5, p < .01$). These results are consistent with differential preparation of the primary and secondary responses, as encouraged by the instructions.

The effects of primary/secondary order relation on secondary-response performance again provides support for a hierarchical coding model. Mean latencies were marginally shorter for hierarchically congruent secondary responses (563 msec) than for incongruent secondary responses (580 msec); $t(18) = 1.8, p < .10$. Durations were significantly shorter for hierarchically congruent secondary responses (617 msec) than for incongruent secondary responses (648 msec); $t(18) = 4.02, p < .01$. Also, error rates were significantly lower for hierarchically congruent secondary responses (5.1%) than for incongruent secondary responses (15.2%); $t(18) = 7.2, p < .01$.

There was also evidence that element-to-position congruence yielded some facilitation for secondary responses. Error rates were significantly lower in the case of position-congruent sequences (8.9%) than in the case of incongruent sequences (15.2%); $t(18) = 7.2, p < .01$. However, latency and

TABLE 5
SYLLABLE SEQUENCES USED IN EXPERIMENT 3

	Secondary sequences		
Primary sequences	Hier. congr.	Element-to-postion congr.	Incongr.
1234	3412	3214	3142
1432	3214	3412	3124
4231	3142	3241	3412
4132	3241	3142	3421
2431	3124	3421	3214
2134	3421	3124	3241

Note. 1 = BEE (/bi/); 2 = BAY (/be/); 3 = BAH (/ba/); 4 = BOO (/bu/).

TABLE 6
RESULTS OF EXPERIMENT 3

	Primary/secondary relationship		
	Hier. congr.	Element-to-position congr.	Incongr.
Latency (ms)			
Secondary	563	576	580
Primary	311	313	311
Duration (ms)			
Secondary	617	639	648
Primary	541	543	544
Error rate (%)			
Secondary	5.1	8.9	15.2
Primary	4.4	2.8	4.8

duration measures revealed no significant differences between secondary responses with position congruence and incongruent secondary responses. Performance was clearly superior for the hierarchically congruent secondary responses than for the position-congruent secondary responses. Durations were shorter in the case of the former sequences than in the case of the latter ($t(18) = 2.9$, $p < .05$) and error rates were lower ($t(18) = 2.7$, $p < .05$).

A breakdown of secondary-error types revealed that of 211 errors, 68.7% were order errors. This percentage did not vary significantly as a function of primary/secondary order relation; $\chi^2 (2) = 1.2$, $p > .10$. About 52.4% of the order errors involved exchanges of positions three and four, and 32.4% involved exchanges of positions two and three. The pattern was similar to that of Experiment 2. There were no significant effects of primary/secondary order relation on primary response performance.

Discussion

The results of Experiment 3 further support a hierarchical coding model, as indicated by the superior performance on hierarchically congruent secondary sequences compared to the incongruent secondary sequences. The experiment also showed some facilitation for the secondary sequences with position congruence. This result could mean that an element-to-position coding model correctly describes some aspect of the production process. However, an explanation of the position-congruence effect in terms of the hierarchical model is also possible. The secondary syllable sequences with element-to-position congruence (i.e., $S_3S_2S_1S_4$) may have caused less disruption to the binary tree structure shown in Figures 1 and 2 than did the incongruent syllable sequences. In particular, the position-congruent secondary sequences maintained a consistent assignment of syllables to left-hand and right-hand branches of the binary tree for the primary sequences. Thus, facilitation due to position congruence can be explained by assuming that restructuring the primary sequence into a binary tree for the secondary sequence took advantage of this overlap. There is no apparent way that the element-to-position coding model can account for the facilitation due to hierarchical congruence.

The results of Experiment 3 also have implications for a model of serial order representation based on element-to-element associations. In particular, such a model has difficulty accounting for the facilitation observed for secondary responses with element-to-position congruence because these responses maintain none of the element-to-

element associations from the associated primary responses. This reinforces the conclusion of Experiment 2 that element-to-element coding is insufficient as a sole account of serial-order representation in the present task.

GENERAL DISCUSSION

Alternative Formulations of the Models

While the results of Experiments 1 and 3 seem to support a hierarchical coding model, it could be claimed that the specific models rejected here do not do justice to the classes of models from which they are drawn. Perhaps generalized versions of the element-to-position or element-to-element coding models provide greater competition for the hierarchical coding model.

In a generalized element-to-position coding model, each element could be connected to all of the positions, with the correct position having the maximum strength of association, and the association strength for other positions being a decreasing function of distance to the correct position. This account would incorporate some relations at a distance, as does the hierarchical coding model. However, it still would not explain our results. The syllables in the present hierarchically congruent sequences (i.e., $S_3S_4S_1S_2$) are farther on the average from their positions in the primary sequence than they were in the secondary sequences with element-to-position congruence ($S_3S_2S_1S_4$). Thus, according to a generalized position-coding model, greater adjustment in association strength would have to occur for the hierarchically congruent sequences than for the sequences with element-to-position congruence, yielding an incorrect prediction about relative ease of performance for the two sequence types (cf. Experiment 3). Analysis of the sequences used in the other experiments also fails to support a generalized element-to-position coding model.

A generalized element-to-element coding model would incorporate associations between nonadjacent elements, with the strength of these associations decreasing as a function of the number of intervening elements. In formulating this model, one must decide how to handle the reversal of direction for some associations when creating a secondary sequence out of a primary sequence. We assume that associations are only useful in specifying order information if they retain their original direction. This implies that sequences with element-to-element congruence ($S_2S_3S_4S_1$) should yield better performance than do hierarchically congruent sequences, since they maintain the association between the syllables S_2 and S_4. Hierarchically congruent sequences do not maintain that association. However, the preceding implication was not supported by Experiment 2, which revealed the opposite relation in ease of performance for these two sequence types. Thus, a generalized element-to-element coding model, like a generalized element-to-position coding model, does not account for the data as well as the hierarchical coding model.

The marginal intermediate results obtained in Experiment 2 for secondary responses with element-to-element congruence are problematic for the hierarchical coding model. One possible explanation of these results is that the response-priming procedure does not isolate a single phase of speech production, and that hierarchical coding is used to represent serial order during one phase, while element-to-element coding is employed during another phase. However, it does seem clear that element-to-element coding models are inadequate as sole accounts of serial-order representation in the present task.

The possibility that the response-priming procedure does not isolate a single phase of speech production suggests that a hybrid model, using different types of serial-order coding, may be necessary in a complete ac-

count of the present results. In addition to accounting for the present results, a hybrid model could perhaps be developed that is consistent with the latency and duration results of Sternberg et al. (1978), which seem most consistent with some form of linear representation of serial order. The present studies indicate that a model of speech production must also account for the finding that speakers impose a hierarchical structure on the production of syllable sequences, even when the sequences themselves lack a hierarchical organization, such as might be specified by prior groupings into words, or by a pattern in the phonemes or syllables present.

Bases for Hierarchical Coding in Speech Production

Short-term memory. One possible basis for hierarchical coding in speech production at the syllable level could involve the organization of short-term memory. It has been argued that movement control uses a short-term motor program buffer (Klapp, Greim, & Marshburn, 1983; Sternberg et al., 1978), and the possible role of this buffer in short-term memory tasks has been explored (Baddeley & Hitch, 1974; Monsell, 1984). Wickelgren (1967) found that subgroups of three elements provided the most natural organization of short-term memory. If hierarchical coding in utterance plans depends on short-term memory groupings, then the present finding of grouping into pairs rather than triplets may have resulted from our use of shorter syllable sequences than those of Wickelgren (1967).

There is some reason, however, to be skeptical about whether our results are due to short-term memory organization. Estes (1972) summarized the results of numerous short-term memory experiments that studied performance for four-element sequences. Plotting the errors observed in them revealed a bow-shaped serial-position curve where the most frequent error was an exchange between positions two and three. None of our experiments revealed such a pattern of results. In fact, we consistently found fewer two–three exchanges than three–four exchanges. This raises doubts about whether the utterance-plan buffer used in rapid speech production of fully prepared utterances overlaps with short-term memory.

While an utterance plan may not be stored in general short-term memory, the findings of Sternberg et al. (1978) do suggest that a temporary representation does play a role in motor control, both for speech production and for manual keypresses. As discussed previously, certain latency and duration data obtained in their studies suggest that this representation has a linear structure. However, Rosenbaum (1985) has argued that the serial-position effects on inter-keypress intervals for the Sternberg et al. (1978) data suggest a more complex organization. He showed that the binary tree model, developed to account for inter-keypress intervals obtained by Rosenbaum et al. (1983), can also account for the Sternberg et al. (1978) serial-position effects.

Rhythmic organization. The hierarchical structure of the fully prepared utterance plan may stem from rhythmic factors, rather than from characteristics of general short-term memory. Many sequential activities (e.g., speech, music, and dance) have a rhythmic component. Lashley (1951) considered rhythm integral to the control of serial order. The potential benefit of doing so is clear from observing that the ordering of linguistic units is simply the timing of linguistic units, stripped of interval properties. A representation that correctly specifies the rhythm of a set of linguistic units inherently specifies their correct order.

As noted earlier, rhythm is best described by relations between elements in a sequence (Liberman & Prince, 1977; Martin, 1972). The simplest such relationship involves alternation of accented and

unaccented elements (Martin, 1972; Povel, 1981; Sweet, 1875, cited in Selkirk, 1984). Sweet (1875, cited in Selkirk, 1984) has termed this phenomenon the "Principle of Rhythmic Alternation." It states that rhythmic patterns, in language and in music, will almost always have a binary structure. Ternary patterns are rare special cases, and quaternary patterns are decomposed into their binary parts. In the theory of metrical phonology (Liberman & Prince, 1977), this binary structure plays a critical role in determining the stress and rhythm of an utterance.

Stress assignment involves specifying the relative prominence of each pair of branches within a hierarchical tree that describes an utterance's surface syntax and lexical structure. Relative prominence is denoted by labeling one binary branch "strong" (or s) and its partner "weak" (or w). Different rules are used at the lexical and syntactic levels for assigning s and w to left and right branches. Figure 5a shows the metrical tree for a sentence with the relative prominence relationships indicated according to the principles of metrical pho-
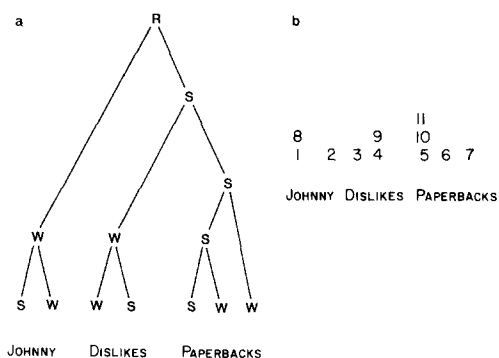

FIG. 5. (a) (After Cruttendon, 1986). A metrical-tree structure for the sentence "Johnny dislikes paperbacks." The "R" indicates the root of the sentence. An s indicates a strong branch, and a w indicates a weak branch. (b) (After Cruttendon, 1986). A metrical-grid structure for the same sentence. The numbers are read off the metrical tree in order, starting at the lowest level of the hierarchy. The numbers of nonterminal nodes are placed over their s daughter nodes.

nology (Liberman & Prince, 1977). The determination of rhythm involves a subsequent stage in which the metrical tree (Figure 5a) is converted to a "metrical grid" (Figure 5b). This representation aligns a parent node with the s element of its pair of daughter nodes. The purpose of the grid is to represent the temporal relationships between stress at different levels of the hierarchy. This formalizes the intuition that there is a tendency to avoid "clashing" stresses. Clashes occur when stresses are in adjacent positions at the same level of the hierarchy, preventing alternation between accented and unaccented elements. In English, some clashes are avoided by stress shifts made according to the "rhythm rule," which causes stresses to appear at more regular temporal intervals.

The theory of metrical phonology, as formulated by Liberman and Prince (1977), shows how a binary tree structure for representing stress can be integrated with other hierarchical influences (syntactic and lexical) as part of language generation. Subsequent developments in metrical phonology (e.g., Selkirk, 1984) have eliminated the metrical tree from the derivation of stress and rhythm, placing greater emphasis on the metrical-grid representation. However, the grid has a hierarchical structure, and binary patterning remains very important there through the Principle of Rhythmic Alternation (Selkirk, 1984).

The present findings suggest that the serial order of rapid fully prepared utterances may have a hierarchical representation with a binary structure. In our experiments, this structure was imposed on the utterance by the speaker, and did not stem from any prespecified pattern in the sequence to be produced, or from prior groupings of the syllables into larger units such as words. Given the generality of the Principle of Rhythmic Alternation, and its usefulness in describing the rhythmic patterns of language, a binary tree representa-

tion of serial-order information may result from general characteristics of rhythmic activity. If this is the case, then the principles of metrical phonology show how the binary tree representation of syllable order in the utterance plan could relate to other hierarchical structures that influence speech production.

REFERENCES

BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

BOCK, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1–47.

BRADSHAW, J. L. (1970). Phonetic homogeneity and articulatory lengthening. *British Journal of Psychology*, 61, 499–507.

CHOMSKY, N. (1963). Formal properties of a grammar. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2). New York: Wiley.

COLLARD, R., & POVEL, D. (1982). Theory of serial pattern production: Tree traversals. *Psychological Review*, 6, 693–707.

CONRAD, R. (1965). Order errors in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4, 161–169.

CRUTTENDON, A. (1986). *Intonation*. 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.

DELL, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283–321.

DELL, G. S., & REICH, P. A. (1980). Toward a unified model of slips of the tongue. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.

DELL, G. S., & REICH, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611–629.

ELLIS, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, 19, 624–634.

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*. Washington, DC: Winston.

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

GORDON, P. C., & MEYER, D. E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 153–178.

GREENO, J. G., & SIMON, H. A. (1974). Processes for sequence production. *Psychological Review*, 81, 187–197.

HENKE, W. (1966). Dynamic articulatory model of speech production using computer simulation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

KEELE, S. W. (1968). Movement control in skilled performance. *Psychological Bulletin*, 70, 378–403.

KEELE, S. W. (1975). The representation of motor programmes. In *Attention and Performance V*. Hillsdale, NJ: Erlbaum.

KELSO, J. A. S., TULLER, B., & HARRIS, K. S. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In P. F. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag.

KENT, R. D., & MINIFIE, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, 5, 115–133.

KLAPP, S. T. (1974). Syllable-dependent pronunciation latencies in number naming: A replication. *Journal of Experimental Psychology*, 102, 1138–1140.

KLAPP, S. T., ABBOTT, J., COFFMAN, K., GREIM, D., SNIDER, R., & YOUNG, F. (1979). Simple and choice reaction time methods in the study of motor programming. *Journal of Motor Behavior*, 11, 91–101.

KLAPP, S. T., GREIM, E. D. M., & MARSHBURN, E. A. (1983). Buffer storage of programmed articulation and articulatory loop: Two names for the same mechanism or two distinct components of short-term memory? In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

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

LIBERMAN, M., & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.

MACKAY, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323–350.

MACKAY, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210–227.

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

MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487–509.

MEYER, D. E., & GORDON, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Language and Memory, 24*, 3–26.

MEYER, D. E., STERNBERG, S., KNOLL, R. L., & WRIGHT, C. E. (1978, May). *Memory retrieval and motor programming of related word sequences.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

MONSELL, S. (1984). Components of working memory underlying verbal skills: A "distributed capacities" view. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X.* Hillsdale, NJ: Erlbaum.

POVEL, D. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 3–18.

ROSENBAUM, D. A. (1985). Motor programming. In H. Heuer, U. Kleinbeck, & K.-M. Schmidt, (Eds.), *Motor behavior: Programming, control, and acquisition.* Berlin: Springer-Verlag.

ROSENBAUM, D. A., INHOFF, A. W. & GORDON, A. M. (1984). Choosing between movement sequences: A hierarchical editor model. *Journal of Experimental Psychology: General, 113*, 372–393.

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

ROSENBAUM, D. A., & KORNBLUM, S. (1982). A priming method for investigating the selection of motor responses. *Acta Psychologica, 51*, 223–243.

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

SCHOURUP, L. (1973). Unique New York unique New York unique New York. In *Papers from the Ninth Regional Meeting, Chicago Linguistic Society.*

SELKIRK, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge: MIT Press.

SHATTUCK-HUFNAGEL, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum.

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. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

STETSON, R. H. (1951). *Motor phonetics.* Amsterdam: North Holland.

WICKELGREN, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *The Quarterly Journal of Experimental Psychology, 19*, 97–102.

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

WICKELGREN, W. A. (1976). Phonetic coding and serial order. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic Press.