

## CONTEXT-SENSITIVE CODING, ASSOCIATIVE MEMORY, AND SERIAL ORDER IN (SPEECH) BEHAVIOR

WAYNE A. WICKELGREN

*Massachusetts Institute of Technology*

The problem of serial order in noncreative behavior is defined in much the same manner as Lashley (1951), and several theories of serial order are examined. Lashley's rejection of associative-chain theories of serial order is shown to apply to one particular theory, and to be invalid as applied to other associative theories. Indeed, the most plausible theory is the "context-sensitive associative theory," which assumes that serial order is encoded by means of associations between context-sensitive elementary motor responses. In speech, this means that a word such as "stop" is assumed to be coded "allophonically" as /#st, st, to, to, op, op#/, rather than being coded phonemically as /s, t, o, p/. This theory handles the pronunciation of single words and even phrases in a certain sense.

Some years ago, Lashley (1951) wrote a very thought-provoking paper on the problems of serial order in (noncreative) behavior, mental planning (priming) of future behavior, syntax in (creative) language behavior, the role (or lack of any role) of sensory feedback in the control of rapid coordinated movement, space coordinate systems, rhythmic action, the interaction of temporal and spatial systems, the constant activity of the nervous system, and the incredible complexity of it all.

Apparently, Lashley considered the understanding of all these aspects of the nervous system to be closely related and describable as the (general) problem of serial order in behavior. Without disputing that there may be some relation between any two aspects of neural functioning, I do not think Lashley has made a convincing case for lumping all these problems into a common one called "serial order in behavior." In any event, the present paper will be concerned with only one of the problems discussed by Lashley (1951), namely, the problem of serial order in noncreative be-

havior, particularly the pronunciation of words.

As we shall see, Lashley (1951) made some implicit assumptions about the coding in the nervous system of noncreative behavior sequences just to be able even to pose the problem of serial order. I will make some of the same assumptions explicitly, but I will question one of Lashley's implicit coding assumptions and propose an alternative. This alternative permits an elegant associative-chain solution to the problem of serial order, whereas Lashley's implicit coding assumption did not permit an associative-chain solution, at least not in the usual sense. Since Lashley was apparently unaware of his coding assumptions, he falsely concluded that he had disproved associative-chain theories of serial order in noncreative behavior sequences. The present paper shows that Lashley's rejection of associative theories was premature.

On the constructive side, the present paper has two closely related goals. First, the paper attempts to advance understanding of the problem of serial order by defining useful concepts, analyzing the assumptions necessary to pose the problem of

<sup>1</sup> This work was supported by Grant MH 08890-04 from the National Institute of Mental Health, United States Public Health Service.

serial order, formulating some alternative theoretical solutions to the problem of serial order, and discussing the plausibility of the various assumptions made either to pose or to solve the problem. Second, the paper discusses the issue of the basic code for speech at the central articulatory level, proposing that the "context-sensitive" allophone is the basic unit in production (and probably also in recognition), rather than the "context-free" phoneme.

### POSING THE PROBLEM

#### *Noncreative Behavior*

To pose the problem of serial order in noncreative behavior, we must first assume the existence of noncreative behavior, that is, behavior which occurs repeatedly in an essentially equivalent manner. I consider examples of such behavior to include the pronunciation of almost all single words and familiar phrases by human beings, all animal "speech" and mimicry, and all elementary skilled motor movements such as throwing a ball, running, walking, jumping, turning one's head, moving one's hand from one location to another, etc.

From a physical point of view, one can assume that there is infinite variation in all behavior, for example, infinite variation in the exact manner in which the hand moves from one location to another and infinite variation in the pairs of possible starting and stopping locations. But this is of no psychological significance. It certainly does not suggest that we should treat all behavior as creative. Without doubt, there is a particular "grain" of representation in each sensory and motor dimension which sets limits on the precision of any motor control system. I shall consider this to impose a finite set of (psychological) equivalence classes on the (physically) infinite set of response sequences.<sup>2</sup>

It seems quite useful to distinguish behavior which is fundamentally creative, though lawful, such as speech at the level of syntax, from behavior which is funda-

mentally noncreative, such as speech at the level of word pronunciation and the other examples cited above. We say the same words over and over again. We rarely say a sentence identical to one said before. This seems to me, as it has seemed to many other people concerned with language (Chomsky, 1964; Chomsky & Miller, 1963), to be perhaps the most fundamental fact about language at the level of syntax. If one feels this to be such a fundamental feature of syntax, one should certainly not expect to account for the serial order of phonemes in words in exactly the same manner as one accounts for the serial order of words (or morphemes) in sentences. Thus, it seems important to distinguish creative and noncreative serially ordered behavior in language and in all other behavior.

Naturally, one would prefer theories of serial order in creative and noncreative behavior to have some assumptions in common. However, it is absurd to claim, for example, that the full apparatus of a generative grammar is necessary for serial order in noncreative behavior such as the pronunciation of a single word. If this were so, parrots would have syntactically structured speech. There must be some very significant advance in the human nervous system that permits syntax, which is not necessary for noncreative, skilled motor movements. It only confuses matters to fuzz these two problems together under the common rubric of serial order in behavior.

#### *Single Sequence of Elementary Motor Responses*

To pose the problem of serial order in the manner of Lashley (1951), it is necessary to assume that a noncreative behavior sequence is controlled by a single sequence of internal representatives of elementary motor responses, at the central articulatory level controlling the behavior. To use Lashley's (1951) example, pronunciation of the word "right" is analyzed, as the sequential activation of the central representatives of the "phonemes" /r/, /ī/, /t/

<sup>2</sup> This has nothing to do with whether one uses continuous or discrete mathematics.

or /r/, /a/, /y/, /t/ in that order. This segmentation of a word into phonemes is certainly the conventional analysis used by linguists, and it is reasonable to assume that if speech at the central articulatory level is properly analyzed as a single sequence of elementary movements or positions (phonemes), then so is all noncreative (skilled) motor behavior.

However, with the advent of distinctive feature analysis of phonemes (Jakobson, Fant, & Halle, 1952) it has occurred to some researchers in linguistics and speech communication that speech at the word level might be more properly analyzed as several sequences (of the values of the several feature dimensions) running in parallel. Neither viewpoint is established as correct at the present time, and this is not the place to discuss the issue. The present paper will make the single-sequence assumption in agreement with Lashley and almost everyone else. The point of this discussion is that there is an alternative to the single-sequence hypothesis, which is not completely implausible. If the multiple-sequence hypothesis is true, there is still a serial order problem and it is much more difficult to solve than is the serial order problem for the single-sequence hypothesis. Furthermore, there is a coordination problem for the multiple-sequence hypothesis. The apparently greater difficulties of serial order and coordination are arguments against the multiple-sequence hypothesis, but, of course, they do not disprove it.

Before leaving the single-sequence assumption, it is necessary to make clear that the difficulties in segmenting speech into nonoverlapping parts at levels of the peripheral articulatory musculature or acoustic waveform have no direct bearing on the issue of the segmentability of speech at the central level of the nervous system controlling speech production (or recognition either, for that matter). Between the central speech level and the periphery, there are differential delays in the speed with which features at the central level affect features at the peripheral level for production (or vice-versa, for recognition).

### *Context-Free Coding of Elementary Motor Responses*

Lashley (1951) assumed that behavior sequences are composed of a number of elementary motor responses (emrs) which can and do occur in a large variety of orders. These emrs are considered to be the same regardless of the context of other emrs in which they occur. Or, regardless of whether or not the emrs are unchanged by their context, the internal representatives of these emrs are assumed to be identical for all contexts. By analogy to Chomsky's (1963) terminology for a related concept, this assumption is called context-free coding of elementary motor responses.

### ALTERNATIVE THEORETICAL SOLUTIONS

#### *Context-Free Associative Memory*

This is the alternative which Lashley (1951) showed to be inadequate. Lashley argued for the general inadequacy of this theory by using word pronunciation as an example of serially ordered behavior and showing that word pronunciation could not be handled by this theory.

Lashley's essential argument is as follows. Words are coded as sequences of phonemes (context-free emrs) in the speech system. There are only a small number of phonemes used in any language (on the order of 50), and these are used in a large variety of orders to form the  $10^5$  to  $10^6$  "words" used by a single individual.<sup>3</sup> Thus, as Lashley (1951) correctly argued, the pairwise associations between phoneme representatives would be of little value in providing information concerning the ordering of the phonemes in any particular word.

Similarly, nonverbal emrs appear in many different behavior sequences in a wide variety of pairwise orders. Thus, Lashley argued, the ordering of emrs for a behavior sequence must come from some outside source, not from associations be-

<sup>3</sup> The figure  $10^5$  to  $10^6$  counts all variants of the same basic stem as different and also counts all familiar phrases. Since I know of no tabulation of the number of such "words" used by one individual,  $10^5$  to  $10^6$  is a wild guess.

tween the internal representatives of the emrs themselves.

Actually, the selection of the *unordered* set of phonemes in any given word (in general, the selection of any unordered set of emrs) is not a difficult problem for an associative memory. To see this, it is necessary to say a little more about the situations in which we pronounce words and to make some very general assumptions about the process of pronouncing words.

Imagine that the task is to pronounce the word just seen or heard. The visual or auditory representation of the word at some level of the nervous system must be capable of activating (directly or indirectly) an articulatory representation of the word. With an associative system, one would say that the visual or auditory internal representative of the word was associated (directly or indirectly) to the articulatory representative. If the coding of a word in the visual or auditory or some abstract (verbal) conceptual system is "distinct enough" from the coding of all other words, then the strengths of associations from this word-representative to the phoneme-representatives in the articulatory system provides adequate information concerning the unordered set of phonemes in the pronunciation of any given word. That is to say, the phoneme-representatives of the word are the ones to which the word-representative is strongly associated. Incidentally, with around  $10^{10}$  neurons in the human nervous system, there are more than enough neurons available to have at least one for every word, and indeed for every meaning of every word. So the coding of words can be "distinct enough."

So, within an associative system, we may assume that the unordered set of phoneme-representatives for a particular word is selected by strength of association to the word representative. This means that to obtain the correct ordered set of phonemes one need only worry about competing associations between pairs of phonemes both of which are in the word. This certainly helps, but it does not permit an associative memory with context-free (in this case,

phonemic) coding of emrs to solve the problem of serial order.

Probably the quickest way to demonstrate this is to consider (as Lashley did) the case of phonemic anagrams, such as "struck" (/struk/) and "crust" (/krust/). Obviously, the phoneme-to-phoneme associations cannot be the basis for the serial ordering of the phonemes in these two words, because the unordered sets of phonemes are identical in both cases and the ordering is different.

We need not even get into arguments regarding more sophisticated associative theories, such as whether the summed (direct and remote) digram frequency of a set of phonemes arranged in the correct order for a word is always higher than the summed digram frequency for some incorrect ordering of the same phonemes. Two different orderings of the phonemes which are correct for different words cannot both have the highest summed digram frequency. Even the most powerful use of the information concerning the strengths of associations between phoneme representatives will not handle the problem of the serial order of phonemes in words.

In a similar vein, Lashley pointed out that one can execute some (all?) behavior sequences in backward as well as forward order. Almost any other order is possible also. Associations between adjacent and remote context-free emrs can hardly be used to account for this.

This is Lashley's (1951) definitive argument against a context-free associative memory. Two other arguments given by Lashley in the same article are sometimes cited (e.g., Lenneberg, 1967, pp. 98-99) as evidence against an associative theory of word pronunciation. Neither of these other two arguments carries any weight against associative theories such as the ones discussed in the present paper.

The first argument is that there is not enough time for the auditory or kinesthetic feedback from one phoneme to trigger the next phoneme. There is currently some debate about this, but, in any event, as Lenneberg (1967) agrees, this is not relevant to a purely central associative theory,

which are the only associative theories considered in this paper.

The second argument is that the pronunciation of a phoneme is influenced by the phonemes to be pronounced later. How can such articulatory anticipation be accounted for by an associative-chain theory? The context-sensitive associative theory discussed in the next section handles this in what I regard as the most direct and plausible manner. But even the context-free associative theory can probably handle the effect, and in precisely the manner suggested by Lashley for handling anticipatory errors in pronunciation, namely, by assuming that the unordered set of phoneme representatives is "primed" (partially activated) before beginning to fully activate any single phoneme representative. Since we have already assumed this, nothing has been added to the theory. This can be used to account for anticipatory errors, on the one hand, and anticipatory effects on articulation, on the other hand. The anticipatory articulation effect comes from assuming that partially activated phoneme representatives provide some output to the speech musculature, in addition to the dominant influence of the currently fully activated phoneme representative.

#### *Context-Sensitive Associative Memory*

Use of the term "context-free" to characterize Lashley's assumption regarding the coding of emrs already suggests the alternative assumption, that emrs are context-sensitive (again using Chomsky's terminology by analogy). As we shall see, context-sensitive coding gives us a plausible associative-chain solution of the problem of serial order in noncreative behavior, just what Lashley said was impossible. The solution is not perfect, but it is completely adequate for the pronunciation of words, and undoubtedly for all other kinds of noncreative behavior.

Let  $u, v, w, x, y, z$  stand for emrs in some system, for example, phonemes in the speech system. Let  ${}_y x_z$  stand for the  $x$  which is preceded by  $y$  and followed by  $z$ . Instead of assuming, as Lashley did, that

the internal representatives of  ${}_y x_z$  and  ${}_u x_w$  are identical for all  $x, y, z, u$ , and  $w$ , we shall assume that the internal representatives of  $x$ 's in different local bilateral contexts are different.

Note that, to solve the problem of serial order with an associative memory, all we must assume is that the (central) internal representatives of  ${}_y x_z$  and  ${}_u x_w$  are different; the motor output (emrs) for  ${}_y x_z$  and  ${}_u x_w$  could be identical. However, there is another knotty problem concerning serial motor behavior that can be solved simultaneously with the problem of serial ordering, if we assume that the emrs for  ${}_y x_z$  and  ${}_u x_w$  can be (and often are) different also. The knotty problem is that the motor neuron activation pattern necessary to cause some part of the body to assume a particular position is different for different prior positions of that part of the body. This is obviously no problem, if the emrs as well as their (central) internal representatives are considered to be context-sensitive.

Of course, we must consider  ${}_y x_z$  and  ${}_u x_w$  to be more similar than  ${}_y x_z$  and  ${}_u v_w$ . Otherwise, we could not have come to analyze behavior into  $/y/, /x/, /z/$  sequences, which could then be more finely analyzed into  $/.y_x/, /{}_y x_z/, /{}_x z./$  sequences. Thus, we must consider the set of elements  $\{{}_y x_z\} = .x$  for a particular  $x$  and all  $y$  and  $z$ , to form a "similarity class."<sup>4</sup>

For speech, as an example, this is all quite reasonable. Although the evidence is only just beginning to be acquired, it appears that the motor neuron activity is quite similar for the different allophones of a

<sup>4</sup> The context-sensitive emr is an alternative to the "gamma loop" (proprioceptive feedback) system, which also permits "go to" motor commands that can achieve a particular position, regardless of the prior position. The context-sensitive emr has two advantages over the gamma loop system for rapid, coordinated movement. First, it does not require feedback, which is advantageous so long as the load on the muscles is predictable in advance. Second, it can achieve a position which is not exactly identical regardless of context. Instead, the position can be in a similarity class of positions that are functionally equivalent, with the particular position being the member of the similarity class that permits the fastest transition to and from it in the given context.

phoneme (Fromkin, 1966; Harris, Ly-saught, & Schvey, 1965; MacNeilage, 1963), but systematic differences are also found (Fromkin, 1966; Harris, 1963; Harris, Huntington, & Sholes, 1966; MacNeilage & DeClerk, 1967). See Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) for a discussion of coarticulation effects at the levels of the physical acoustic signal, the vocal tract configuration, and the locally summated motor neuron activity (electromyogram—EMG). Their discussion emphasizes the relatively greater similarity among the allophones of a phoneme at the level of motor neuron activity than at either of the other two levels (especially the acoustic signal). Nevertheless, Liberman et al. acknowledge that evidence exists for coarticulation effects, even at the motor level. Incidentally, coarticulation effects at the vocal tract level rule out the exclusive use of “go to” gamma loop commands (MacNeilage & DeClerk, 1967).

To see just how the assumption of context-sensitive emrs solves the problem of serial order, let us consider again the phonemic anagrams, /struk/ and /krust/. Written in terms of context-sensitive emrs (context-sensitive allophones<sup>6</sup>), the unordered sets for these two words are no longer identical. They are /#s<sub>t</sub>, s<sub>t</sub>r, t<sub>r</sub>u, r<sub>u</sub>k, u<sub>k</sub>#/ and /#k<sub>r</sub>, k<sub>r</sub>u, r<sub>u</sub>s, u<sub>s</sub>t, s<sub>t</sub>#/. Assuming again that the unordered set of context-sensitive emr representatives is partially activated by the word representative, then all that is needed to achieve full activation of the set in the correct order is the well-learned association from the internal representative of “begin” to the set of emr representatives #x<sub>s</sub>. In the case of /struk/, this will cause #s<sub>t</sub> to be the most strongly activated emr, at first. Then

s<sub>t</sub>r will be most strongly associated to #s<sub>t</sub> among all the partially activated emrs composing the word, so s<sub>t</sub>r will be the next to be most strongly activated, and so on. Note that we are implicitly assuming that an emr is emitted whenever its internal representative is the most strongly (fully) activated, and that, once fully activated, an emr-representative only remains fully activated for a short period of time before being inhibited and transferring control to the next emr-representative. The partial activation of each member of the unordered set of emr-representatives obviously must persist for the time until that member is fully activated.

This theory handles the basic problem of serial order for all noncreative behavior sequences which do not involve two or more identical pairs of emrs followed by a different emr. For example, this theory will not, by itself, account for the ability to pronounce /lampblak/ correctly. This sequence has two “identical” pairs of adjacent phonemes followed by a different phoneme in the two cases: /la/ followed by /m/ and later /la/ followed by /k/.

Now, of course, humans are able to pronounce /lampblak/. Does this show that the context-sensitive associative theory is wrong? Definitely not. In the first place, there are extremely few instances in English of repetitions of a phoneme pair followed by different phonemes. Examination of the 3,800 words beginning with b, d, f, and l and occurring at least once in 10<sup>6</sup> words according to the Thorndike-Lorge (1944) count uncovered exactly 12 cases where a phoneme pair was repeated followed by a different phoneme. The words were: barnyard, brethern, fair-haired, farmyard, foreshorten, forlorn, fourscore, lampblack, Lapland, lifelike, limelight, and lullaby.

One could handle these cases by assuming that these words were pronounced in the correct order by breaking each word into two parts, each part being handled by the method described in this section. The ordering of the parts is considered as part of the unsolved problem of the grammatical ordering of words in sentences. This res-

<sup>6</sup> “Allophone” is used in this paper to stand for a phoneme in a particular context of phonemes on either side. Most specifically, “allophone” stands for a phoneme with one phoneme specified before and after it. In other words, the allophone concept is defined in terms of the phoneme concept, and refers to a class of similar speech sounds or gestures occurring in a specified environment. This is closely related to, but somewhat different from, the use of the term “allophone” in linguistics.

olution of the problem seems intuitively unsatisfying, and, fortunately, there is another solution.

The more satisfying solution derives once again from questioning the phonemic coding assumption. Since the present theory assumes that the coding of words is in terms of context-sensitive allophones, not phonemes, there is no reason not to consider stress as a feature distinguishing among different emrs for speech. Thus, the two /a/s in /lampblak/ are not identical; they differ in stress.

Having made the assumptions that the internal representatives of phonemes in different local bilateral contexts or with different stress are different, we have solved the problem of serial order in word pronunciation, with an associative memory. Furthermore, as will be shown in a later section, this associative solution can be extended to handle the serial order of verbal emrs in novel (or familiar) phrases consisting of several words.

It should be noted that context-sensitive coding of speech is not equivalent to syllabic (CV, VC, and/or CVC) coding. There are certain similarities. Both predict coarticulation effects at the motor neuron level within syllables. However, the context-sensitive coding hypothesis predicts coarticulation effects across syllable boundaries, which a syllabic coding hypothesis would not predict. To my knowledge, no tests of coarticulation effects at the motor neuron level have been performed with multisyllabic words. Of course, one must try out all versions of the syllabic coding hypothesis, CV, VC, CVC, or a specified mixture of these. However, the context-sensitive coding hypothesis predicts that coarticulation effects can occur anywhere, because context-sensitive coding is essentially equivalent to coding in terms of overlapping syllables. So any nonoverlapping syllabic coding hypothesis can, in principle, be distinguished from context-sensitive coding. Furthermore, syllabic coding is a context-free (alphabetic or phonemic) code, and thus it cannot provide information on the serial ordering of syllables in a word. Hence, syllabic coding does not solve the

problem of serial order between syllables. Furthermore, it does not solve the problem of the serial ordering of the components of the syllable, the emrs. Context-sensitive coding handles both serial order problems by the same mechanism.

Before leaving this theory, it is necessary to consider whether we have assumed too many internal representatives of emrs in the speech system compared to the number of neurons in the human brain. Assuming that stress is a feature of vowels and that there are four different levels of stress, we have about 50 different English vowels. This, combined with the 24 English consonants, makes a total of about 75 basic similarity classes of emrs; call it 100 to be on the safe side, as we might want to consider certain consonant clusters as emrs at the most central articulatory level. If all triples of the basic similarity classes of vocal emrs occurred in English (which they do not), then  $10^2 \times 10^2 \times 10^2 = 10^6$  internal representatives would be necessary in order to encode all the context-sensitive emrs in English speech. This is small in comparison to  $10^{10}$ , so we need not worry about not having enough neurons to code speech in this manner. In fact, the large number of neurons in the human nervous system suggests that the nervous system generally uses a very large number of elements as a basic feature of its approach to control systems problems. The context-sensitive associative theory of serial order is in line with this general principle.

#### *Contingent Associative Memory*

For this theory we can return to the assumption that a word is coded as an ordered set of phonemes (context-free emrs). The word representative of "crust," for example, is still associated to the unordered set of phoneme representatives /k/, /r/, /u/, /s/, /t/, and to some extent, the order of the phonemes is still determined by associations between the phoneme representatives. However, a very powerful kind of interaction is assumed to exist between a word representative and the associations between phoneme representatives, namely, the capacity of a word repre-

sentative to potentiate the correct (direct, forward) associations between the phoneme representatives of the word and/or to inhibit the incorrect (remote and backward) associations. For example, activation of the word-representative "crust" potentiates the associations between /k/ and /r/, between /r/ and /u/, between /u/ and /s/, and between /s/ and /t/, in addition to partially activating the correct set of phoneme representatives.

Contingent association also solves the basic problem of serial order in behavior, though there are some remaining difficulties with words involving repeated occurrences of the same phoneme. These difficulties can be solved by assuming that remote associations are potentiated in inverse proportion to their degree of remoteness and/or inhibited in direct proportion to their degree of remoteness. There should still be a greater tendency for errors in serial ordering to occur after phonemes that are repeated in a word, just as there should be this differential tendency for the previous associative theory with context-sensitive coding of emrs. Rather than being a defect of the two associative theories, this may be an argument in their favor, if it can be rigorously established that transpositions of phonemes (phoneme-Spoonerisms) occur with higher probability in connection with words or phrases having repeated phonemes than with words or phrases having no repeated phonemes.

#### *Multiple Associative Memory<sup>6</sup>*

According to this theory, a word is coded as an ordered set of phonemes, but the phoneme representative for a given word is completely different from the representative of the same phoneme in a different word. This requires about 10 times as many phoneme representatives as word representatives, but that is not an excessive number in comparison to the number of neurons in the nervous system. Naturally, with this system the ordering of the phonemes in a word can be given by as-

<sup>6</sup> The basic idea for this theory was suggested by Jerry Fodor, but he should not be considered to support this theory.

sociations between the phoneme representatives.

A very similar theory assumes that associated with every word is a set of serial-position representatives standing for the first, second, third, etc. phoneme in the word. Each serial-position representative is associated to the phoneme representative appropriate for that position of that word. This requires exactly the same number of serial position representatives as the number of phoneme representatives required by the former version of the theory, and the two theories seem behaviorally indistinguishable.

#### *Nonassociative Memory*

Nonassociative memories (such as tape recorders, buffer storage in a computer, or after-images on the retina) use an ordered set of locations (registers, cells, boxes, etc.) into each of which can be encoded any member of the relevant set of internal representatives. In our example of the pronunciation of words, this means an ordered set of locations into each of which can be encoded the internal representative of any phoneme. There must be at least as many locations in this nonassociative buffer store as there are phonemes in the longest word or phrase that is assumed to be pronounced as a single "unit" (in a single step). A word representative is assumed to activate the representative of its first phoneme in the first location of the memory, the representative of its second phoneme in the second location of the memory, etc. Read-in of the proper phonemes for the word may be assumed to be accomplished either simultaneously or successively, but read-out (to determine the noncreative behavior sequence) must be assumed to occur sequentially, with the order given by the fixed ordering of the locations.

A nonassociative memory of this type seems to be just the sort of memory Lashley (1951) wanted, since the ordering of the phonemes "is imposed by some other agent" than associations between the representatives of the phonemes composing the word. Because nonassociative mem-



ories make no use of associations between the internal representatives inside the locations, they have trouble accounting for why sequences with repeated elements (in this case, phonemes) should be handled any differently than sequences without repeated elements. To my knowledge, there is very little definite evidence for differential error rates in the pronunciation of words or phrases with repeated phonemes as opposed to those without repeated phonemes. However, my personal experience with phoneme-Spoonerisms in spontaneous speech suggests effects of repeated phonemes, and such effects have been found in verbal short-term memory (Wickelgren, 1965, 1966) and the writing of a disgraphic (LeCours, 1966). Wherever effects of repeated elements are found, nonassociative memories are very improbable. Incidentally, the previous multiple-associative memory would also have trouble handling any repeated-item phenomena.

#### EVALUATION OF THE THEORIES OF SERIAL ORDER

After defining the problem of serial order in noncreative behavior and showing exactly what kind of associative theory is inadequate to handle it, four alternative theories were proposed, each of which can handle the basic phenomenon of serial order in noncreative behavior. The pronunciation of words has been used as an example of noncreative serially-ordered behavior. In accordance with Lashley (1951), who followed the same plan, I consider the problem of serial order in noncreative speech behavior to be identical to the problem of serial order in any other kind of noncreative behavior. However, one must recognize the possibility that the solution to the problem could be different in different cases, unlikely as that might seem.

The four theories can be thought of as ordered on an associative—nonassociative continuum, with the context-sensitive associative theory being clearly in the domain of what has been considered in the past to be an associative theory, the contingent associative theory being a new and more power-

ful kind of associative theory, and the multiple associative theory being similar in many respects to a nonassociative theory. Since all four theories solve the basic problem of serial order, it is clear that Lashley's (1951) rejection of associative solutions was premature. There are many ways to solve the problem of serial order in noncreative behavior, and some of these ways are associative to a greater or lesser extent. Is there any way, at present, to decide which of these theories is most likely to be correct for human beings?

#### *Repeated-Item Phenomena*

One way, which has been mentioned already, is to look for the kinds of repeated-item phenomena predicted by either of the first two associative theories. These phenomena would not be expected if either the nonassociative or the multiple associative theory were correct. Repeated-item phenomena have proved very useful in demonstrating that verbal short-term memory is associative (see Wickelgren, 1965 and 1966 for the findings and a detailed presentation of the arguments), and there is some reason to think that similar phenomena would be found for noncreative speech behavior and other forms of skilled motor behavior. In fact, such repeated-item phenomena already have been found in the writing of a disgraphic (LeCours, 1966).

#### *Difficulty or Impossibility of Contingent Associative Learning*

The extraordinary difficulty human beings have in establishing truly contingent associations in a rote-learning task (Chang & Shepard, 1964) is one argument against the contingent associative theory. Truly contingent associations are ones where each element of a compound stimulus item provides no information concerning the correct response, because it is equally often paired with each possible response item. Only the combination of two or more stimulus elements tells you which response to choose. For example, in one experiment Chang and Shepard employed the eight stimulus items, BAG, BAT, BUG, BUT, RAG, RAT, RUG, and RUT, paired with two response items in such a

way that each stimulus letter was paired equally often with each response. Chang and Shepard found contingent associative learning vastly more difficult than other kinds of classification learning.

Many of Chang and Shepard's subjects were able to learn what appear to be contingent associations. However, so far we have implicitly assumed context-free (phonemic) coding. If we assume context-sensitive (allophonic) coding, then the internal representative of each stimulus element in the Chang and Shepard experiment was at least somewhat different depending on which elements it was adjacent to. Thus, according to the context-sensitive coding theory, this was not truly contingent associative learning, (though it undoubtedly approached it), and a context-sensitive, but noncontingent, associative memory could have mastered the Chang and Shepard task eventually.

One can formulate a truly contingent associative learning task within the context of the context-sensitive coding hypothesis. What is required is to separate the relevant stimulus elements by elements that are identical for every stimulus. Provided subjects learn by rote, this procedure insures that the task requires contingent associations, even if humans use context-sensitive coding. This is so because the context of the relevant elements is identical in every stimulus. However, the task must be learned by rote, since there are a variety of learning strategies that would avoid the necessity of establishing contingent associations. Suppressing the identical (irrelevant) stimulus elements is one such strategy. Using visual memory is probably another. The fact that these learning strategies are, at present, outside the scope of any of our memory theories should not blind us to their existence. Thus, it seems relatively fruitless to perform experiments to determine whether or not subjects can form contingent associations, unless we are fairly confident that we can eliminate all learning strategies that would get around the need to form contingent associations.

I ran myself in a contingent association task which was virtually guaranteed to

require that contingent associations be formed in order to learn the list. The following procedures were adopted: (a) The 16 stimuli consisted of all the combinations of D, V, X, or J in the first position, and +, :, ?, or % in the third position, with the word PLANK always in the second position. (b) The 4 response items (peach, pear, prune, and apple) were equally often paired with each stimulus element. (c) Each compound stimulus was extremely difficult to interpret as a single meaningful concept on the basis of past experience. This made it easier to avoid suppressing the irrelevant word PLANK. (d) I attempted to learn completely by rote (pronouncing each correct pair four times after seeing the answer). I employed no strategies, such as omitting the middle stimulus element, visualizing lists or matrices, learning selected pairs and figuring out the answers to the rest of the pairs through logical reasoning involving knowledge of the way the list was constructed, etc.

The first time I tried to learn this list of contingent associations, I believed in contingent association and expected to be able to demonstrate it. After seven trials over a 3-day period (to allow time for consolidation) with no progress whatsoever, I became extremely discouraged and quit. Five months later, when I no longer believed in contingent association, I decided that more trials would be necessary to convince anybody else, so I ran myself eight trials a day for 4 days and again made no progress whatsoever. Of course, even if these informal results are replicated by other people, it will only disprove contingent association for temporally very near, but nonadjacent, stimulus components. At present, I cannot see how to distinguish experimentally between adjacent contingent association, on the one hand, and context-sensitive coding with ordinary noncontingent association, on the other hand.

However, to establish the necessary contingent associations for pronunciation of words, according to the contingent associative theory, it is necessary to assume that contingent associations be able to be formed involving a word (concept) representative

and a *number* of phoneme representatives. Only one of the phoneme representatives could actually be activated immediately after the word representative was activated, so the rest must be "temporally near but not adjacent." Thus, the extreme difficulty or impossibility of contingent-association tasks casts doubt on the contingent associative theory of serial order.

### *Coarticulation Effects*

Further evidence in favor of the context-sensitive associative theory of the pronunciation of words comes from coarticulation effects. By assuming (context-sensitive) allophones to be the basic unit of articulation, rather than (context-free) phonemes, it is trivial to account for how the "same phoneme" in different phonemic environments can be and must be different in some respects at all levels of the speech process, including the acoustic, vocal tract, and motor neuron levels. If achieving the same vocal tract configuration from different starting positions requires different motor neuron activity (principally quantitative differences in intensity and duration), then it seems quite reasonable to assume that the (context-sensitive) allophone, rather than the (context-free) phoneme, is the basic unit of articulation.

From the present point of view, coarticulation effects are not some strange complication to be disposed of by searching for a level of the speech process at which they do not exist, but are, instead, a basic feature of the speech code at all levels. At the most central articulatory level, context-sensitive coding solves the problem of serial order. At the peripheral motor level, context-sensitive coding solves the problem of achieving a similar terminal position from different starting positions. Many of the problems of speech recognition are also eliminated by assuming (context-sensitive) allophonic coding, since the acoustic properties of an allophone, with one preceding and one succeeding phoneme (allophone class) specified, are relatively invariant.

It seems very difficult for any theory specifying context-free coding to account for coarticulation effects at any neural level,

and, in fact, it is the current hypothesis of certain researchers in speech (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) that coarticulation effects occur only in the transformations from vocal tract configurations to acoustic signal and from motor neuron activity to vocal tract configuration. While the evidence that exists at present appears to show greater *similarity* between the allophones of a phoneme at the motor neuron level than at the vocal tract and acoustic levels, it does not appear to show *invariance*. Since coarticulation effects appear to be found at the motor neuron level, context-sensitive (allophonic) coding is likely for the more central levels of the articulatory system as well.

What is the status of the phoneme by this account? It is an allophone class, undoubtedly based on similarity of articulation, which as Liberman et al. have emphasized is even greater at the neural levels than it is at the level of vocal tract configuration. Also, we have specified the context to which emrs are sensitive in terms of phonemes (allophone classes) rather than allophones. After all, we must avoid an infinite regress at some point, and we cannot make our emrs sensitive to too much context on either side or we will run out of neurons. Put differently, the definition of context-sensitive emrs requires the definition of emr classes. So if context-sensitive allophones are represented in the nervous system, then context-free phonemes must also be represented. According to the context-sensitive associative theory, phoneme representatives need not play a role in the control of articulation, but they must be quite important for other purposes, such as learning to speak.

### *Pronunciation of Phrases*

Finally, let us assume that a human being plans an entire phrase of many words and then articulates the entire phrase as an automatic process, even though that phrase has never before been articulated by him or heard by him. If we make that assumption, then the context-sensitive associative theory and the multiple associative theory

seems to have a definite advantage over the others for this extension of the pronunciation process from words (and familiar phrases) to unfamiliar phrases.

We cannot assume there to be a phrase representative, different from the representatives of the words in the phrase, for every phrase we utter. There are too many phrases in comparison to the number of neurons in the brain. So we shall assume that the internal representation of a phrase is the ordered activation of the internal representatives of the words composing the phrase. We are not concerned, in the present paper, with how the word representatives came to be activated in grammatical, meaningful order. These are the problems of syntax and semantics, for which, *at present*, no associative theory gives an adequate account.

We are concerned with how an ordered activation of word representatives produces an ordered activation of vocal articulatory representatives (phoneme or allophone representatives). Pronunciation of a phrase is assumed to require more time than the planning of the phrase (selection of the ordered set of word representatives), which is reasonable since pronunciation involves mechanical movement. It would be highly desirable to plan the phrase as a whole and input the information regarding the phrase to the central articulatory system in a rapid manner that was complete before the actual control of pronunciation by the articulatory system began. Then during the (slow) pronunciation process, the conceptual system could be planning what to say next. Besides being efficient, this is in accord with intuition.

*Context-sensitive associative theory.* Two phases, input and output, must be distinguished. The input (or priming) phase refers to the period during which the ordered set of words is activated at the conceptual level, partially activating, in order, the unordered sets of allophone representatives corresponding to each word at the articulatory level. In addition to representatives of the words of the phrase, there are assumed to be unpronounced representatives of "beginning" and "end" of the phrase at

both conceptual and articulatory levels. Also during input phase, by short-term contiguity conditioning, "begin" is assumed to be most strongly associated to the set of allophone representatives for the first word, the set of allophone representatives of the first word are most strongly associated to each other, but next most strongly associated to the allophone representatives of the immediately following word, and so on to the "end."

Thus, input phase does two things. First, it partially activates (primes) the sets of allophone representatives for all the words in the phrase. Thus, only these allophone representatives are likely to become fully activated during pronunciation of the phrase (output phase). Considering the very large number of allophone representatives (probably on the order of  $10^4$  to  $10^5$ ), selection of 100 or so for a medium-length phrase is quite a significant accomplishment, informationally. Second, it establishes the ordering of the sets of allophone representatives by means of a gradient of short-term memory traces (greatest strength of association within the allophone representatives of a word, next greatest to the allophone representatives of the next word, etc.).

During output phase, the concept "begin" is always fully activated first. Because of the short-term associations, "begin" most strongly activates the set of allophone representatives for the first word. Because of the long-term associations, "begin" activates \*x, representatives more strongly than any other allophone representatives. Thus, the representative of the first allophone of the first word will be most strongly activated in the beginning. After the first allophone representative has been activated for a duration set by a separate speech timing mechanism, it will be inhibited, and the allophone representative which has the greatest degree of activation will be the next to be fully activated (and determine pronunciation). This will almost always be a representative of another allophone in the first word, because all the representatives of the allophones of the first word are somewhat more strongly

activated than those of subsequent words due to input from "begin" (and short-term associations to each other which tend to maintain this heightened degree of activation even after "begin" was inhibited). Furthermore, because of the long-term associations between the context-sensitive allophones, the representative of the second allophone of the first word will almost always be the one activated after the representative of the first allophone of the first word, and so on to the end of the word. When the first word is finished, the allophone representatives of the second word will be the most highly activated due to the short-term associations from the allophone representatives of the first word, and the representative of the first allophone of the second word ( $*x_2$ ) will be most strongly activated because of the long-term associations from the representative of the last allophone of the first word ( $\vee u_1$ ). The output phase continues in the same manner to the "end."

Thus, the context-sensitive associative theory handles the pronunciation of novel phrases at the articulatory level without further intervention by the conceptual level. The key reason why this is possible for the context-sensitive associative theory is that there is an encoding of the transition from one allophone to another at the articulatory level. However, as stated thus far, coarticulation effects should not cross word boundaries, since each word starts with a  $*x_i$  and ends with a  $\vee u_i$ . If coarticulation effects do cross word boundaries, then the theory will have to be modified. One complex suggestion for how to do this is given in a later section.

*Multiple associative memory.* The multiple associative memory encodes transitions between successive phonemes in a word at the phonetic level, and it can encode transitions between words at the phonetic level by short-term contiguity conditioning of the phoneme representatives of one word to the phoneme representatives of the next word. Thus, the multiple associative memory also permits planning of an entire phrase at the conceptual level and then execution of the pronunciation of that

phrase at the articulatory level, without "tying-up" the conceptual level during the pronunciation process. Hence, one can be planning the next utterance at the word level, while controlling the pronunciation of the present utterance at the articulatory level. There does not seem to be any way to get coarticulation effects to cross word boundaries with the multiple associative theory, but this is not evidence against the multiple associative theory unless such effects can be demonstrated at some neural level.

*Contingent associative memory.* With the contingent associative memory, on the other hand, the word-representative must be activated at the time of pronouncing each word, in order even to pronounce the phonemes in the correct order for a single word, let alone to mediate the correct transition from word to word in the phrase. Thus, the conceptual level must be tied-up, at least intermittently, at the beginning of the pronunciation of each word in a phrase.

*Nonassociative memory.* With the nonassociative memory, it seems necessary to fill the nonassociative memory locations with the correct phoneme representatives for a single word at a time. The reason for this is not because a nonassociative memory with a capacity of 100 or 1000 locations is unreasonable, but rather because it is not clear how a word representative could be conditioned to the phoneme representatives in each location in such a way as to activate the correct ordered set of phoneme representatives in an ordered set of locations, starting from any location in the memory. Perhaps there is some simple way the nervous system could accomplish this feat, but I do not see it. Hence, the nonassociative memory does not seem to be able to pronounce a phrase as a unit without word-by-word input from the conceptual level.

### *Coarticulation Effects Crossing Word Boundaries*

Although such effects have not even been investigated at the motor neuron level, to my knowledge, it is of theoretical interest to discuss the problem and describe one

way the context-sensitive associative theory could be modified to handle such effects, were they to be found. I see no simple way to modify any of the other three theories to handle coarticulation effects crossing word boundaries.

The primary modification of the context-sensitive associative theory consists of assuming that, for a word in isolation, the internal representatives of its initial and terminal allophones are not the specific representatives,  $*\delta_{\Lambda}$  and  $\nu_{\Lambda}$ , respectively, but rather the *sets* of allophone representatives,  $\delta_{\Lambda}$  and  $\nu_{\Lambda}$ , respectively, where  $\delta$  stands for any phoneme. That is to say, input from the word representative leaves unspecified the prior phoneme of the initial allophone of a word and the subsequent phoneme of the terminal allophone of a word. To obtain coarticulation effects that cross word boundaries, the specification of the initial and terminal allophone representatives for a word must be completed by the set of allophone representatives partially activated for the prior and subsequent words, respectively. For example, in the phrase, "the fretful elk," which we shall now represent allophonically as  $/\delta_{\Lambda}, \delta_{\Lambda t}, \Lambda f_r, r_e, r_e t, e t, t f_u, t U, u l_e, i e, e l_k, i k_*/$ , the word representatives partially activate almost this entire unordered set. However, instead of the specific allophone representatives,  $*\delta_{\Lambda}, \delta_{\Lambda t}, \Lambda f_r, u l_e, i e$ , and  $i k_*$ , the word representatives even less partially activate the *sets* of allophone representatives,  $\delta_{\Lambda}, \delta_{\Lambda t}, f_r, u l_e, e l_e$ , and  $i k_e$ , respectively.

Now it is desirable to explain how an associative memory could "fill in the blanks," in other words, how the correct allophone representative could be selected (partially activated, instead of less partially activated) in each set. There are probably several ways to do this. One way is the following. Since no word precedes "the," we may assume that the concept "begin" is strongly associated to the set  $*\delta_{\Lambda}$  causing  $*\delta_{\Lambda}$  to be selected (activated) more than any other member of the set  $\delta_{\Lambda}$ . Similarly, at the end of the phrase,  $i k_*$  is selected by association to the cue "end."

The problem arises only in the interior of

the phrase. The solution in the interior is similar to the solution at the ends, with the addition of the following plausible, but ad hoc, assumption. The sum of the degrees of activation of all the representatives in the set  $\delta_{\Lambda}$  or  $\nu_{\Lambda}$  is greater than the degree of activation of any particular  $\delta_{\Lambda}$ . If this is true then in the above example, due to the short-term contiguity conditioning of adjacent sets of allophone representatives,  $\delta_{\Lambda}$  will get more input from  $f_r$  than from any other representative. The representative  $f_r$  will be most strongly associated to  $\delta_{\Lambda t}$  of all the members of the set  $\delta_{\Lambda}$ , so  $\delta_{\Lambda t}$  will come to be partially activated more than every other  $\delta_{\Lambda}$ . Similarly,  $f_r$  will be specified primarily by  $\delta_{\Lambda}$  to become  $\Lambda f_r$ , etc. Thus, the context-sensitive associative theory seems able to handle coarticulation effects crossing word boundaries, should they be found.

#### *Suprasegmental Features (Stress)*

There is one further problem that must be dealt with by the context-sensitive associative theory, the problem of the different stress (intonation, intensity, duration pattern) of a word in different phrases. For an excellent treatment of stress see Lieberman (1967). For present purposes, stress can be handled by assuming that the input from the conceptual level is an ordered activation of word-stress pairs, that is, the representatives of a word and its stress (the complete pattern of suprasegmental features) are activated simultaneously at the conceptual level. The word representative activates the appropriate set of allophone representatives at the articulatory level. The stress representative at the conceptual level activates a stress representative (or a set of context-sensitive stress representatives) at the articulatory level which is strongly associated to the allophone representatives by short-term contiguity conditioning. During the pronunciation of a word the segmental features are set by the allophone representatives and the suprasegmental features temporarily associated with that set of allophones are set by the stress representative (or sequence of stress representatives). Re-

membering that the intrinsic differences in stress of different syllables of the word are coded by the allophone representatives, we have handled both segmental and suprasegmental stress.

Essentially the same solution applies in the three other theories so suprasegmental phenomena appear to be irrelevant to the determination of the correct theory of serial order in noncreative behavior.

### CONCLUSION

Lashley's (1951) alleged demonstration that associative theories cannot account for serial order in noncreative behavior applies to one particular associative theory, which assumes context-free coding of emrs. Other associative theories, as well as the nonassociative theory proposed by Lashley, can account for the basic phenomenon of serially ordered behavior. Present knowledge does not permit a definite decision concerning the correct theory of serial order in noncreative human behavior. At the present time, the context-sensitive associative theory seems more likely to be correct than the contingent associative theory, the multiple associative theory, or the nonassociative theory. Only the context-sensitive associative theory handles repeated-item phenomena, coarticulation effects within words and across word boundaries, the pronunciation of phrases as a unitary process, and the apparent inability of humans to establish truly contingent associations. Each of the three other theories fails to handle one or more of these phenomena. However, on the one hand, most of the phenomena require more evidence in order to be definitely established, and, on the other hand, it might be that a slight modification of one of the three other theories would make it fit all the facts.

### REFERENCES

- CHANG, J.-J., & SHEPARD, R. N. Meaningfulness in classification learning with pronounceable trigrams. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 85-90.
- CHOMSKY, N. Formal properties of grammar. In R. D. Luce, R. Bush, & E. Galanter (Eds.),
- Handbook of mathematical psychology*. Vol. II. New York: Wiley, 1963.
- CHOMSKY, N. *Current issues in linguistic theory*. The Hague: Mouton, 1964.
- CHOMSKY, N., & MILLER, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. II. New York: Wiley, 1963.
- FROMKIN, V. A. Neuro-muscular specification of linguistic units. *Language and Speech*, 1966, 9, 170-199.
- HARRIS, K. S. Behavior of the tongue in the production of some alveolar consonants. *Journal of the Acoustical Society of America*, 1963, 35, 784 (abstract).
- HARRIS, K. S., HUNTINGTON, D. A., & SHOLES, G. N. Coarticulation of some disyllabic utterances measured by electromyographic techniques. *Journal of the Acoustical Society of America*, 1966, 39, 1219. (abstract)
- HARRIS, K. S., LYSAGHT, G., & SCHVEY, M. M. Some aspects of the production of oral and nasal labial stops. *Language and Speech*, 1965, 8, 135-147.
- JACOBSON, R., FANT, C. G. M., & HALLE, M. *Preliminaries to speech analysis*. Cambridge: M. I. T. Press, 1952.
- LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.
- LECOURS, A.-R. Serial order in writing—a study of misspelled words in "developmental dysgraphia." *Neuropsychologia*, 1966, 4, 221-241.
- LENNEBERG, E. H. *Biological foundations of language*. New York: Wiley, 1967.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- LIEBERMAN, P. *Intonation, perception and language*. Cambridge: M.I.T. Press, 1967.
- MACNEILAGE, P. F. Electromyographic and acoustic study of the production of certain final clusters. *Journal of the Acoustical Society of America*, 1963, 35, 461-463.
- MACNEILAGE, P. F., & DECLERK, J. L. On the motor control of coarticulation in CVC monosyllables. Paper presented at the 1967 Conference on Speech Communication and Processing, Cambridge, Mass., November 1967.
- THORNDIKE, E. L., & LORGE, I. *The Teacher's word book of 30,000 words*. New York: Columbia University Bureau of Publications, 1944.
- WICKELGREN, W. A. Short-term memory for repeated and non-repeated items. *The Quarterly Journal of Experimental Psychology*, 1965, 17, 14-25.
- WICKELGREN, W. A. Associative intrusions in short-term recall. *Journal of Experimental Psychology*, 1966, 72, 853-858.

(Received December 7, 1967)