JOURNAL OF MEMORY AND LANGUAGE **35,** 1–31 (1996)
ARTICLE NO. 0001

A Linguistically Constrained Model of Short-Term Memory for Nonwords

TOM HARTLEY AND GEORGE HOUGHTON

University College London, United Kingdom

Errors in the repetition and serial recall of nonwords indicate that structural properties of the syllable are represented in short-term memory. We develop a connectionist model of short-term memory for such unfamiliar phonological sequences, based on insights from existing models of speech production and short-term memory. The results of simulations of experiments involving nonword recall and repetition are presented, and the mechanisms which produce the common error types are discussed. We also show how the proposed model can be extended to develop and make use of long-term representations of phonological forms, and how a global reduction in short-term memory capacity (such as is commonly observed in developmental dyslexia, and acquired disorders of output phonology) could differentially affect performance on tasks involving familiar and unfamiliar materials.

In this paper we present a linguistically constrained model of the learning and recall of unfamiliar words in verbal short-term memory. All the words a mature speaker knows were once new to them, but normal speakers, even very young children, can often repeat a nonword after a single exposure (Gathercole & Baddeley, 1989; Gathercole & Adams, 1993). The apparent simplicity of this task disguises what may be a rather complex system dedicated to the solution of a specific problem—the need to represent and recall serially ordered verbal stimuli. Spoken words are spread over time, so that there is no point at which all of the information to be retained is concurrently present. The serial structure of the stimulus (e.g., the order of phonemes in a syllable) is therefore central to the identity of the stimulus and must be retained. Once spoken, the word is no longer present in the environment, and cannot be reexamined at will (unlike, say, a typical visual stimulus). In order to repeat or rehearse a novel input, a single trial serial-order learning mechanism is needed. This mechanism must track the input in real-time and have produced a representation capable of supporting rehearsal by the time the stimulus finishes. It is proposed in the current work that this remarkable ability underlies the development of more long-term lexical–phonological knowledge. As well as being of interest in its own right, the explication of this capacity is thus central to the understanding of language acquisition.

The model developed here lies at the confluence of a number of theoretical and empirical streams. Theoretically, our starting point is the class of neural network models of serial order known as “competitive queuing” (CQ) models (Houghton, 1990), as applied to verbal short-term memory by Burgess and Hitch (1992). Our principal aim has been to incorporate phonological constraints on syllable structure into these models. The empirical motivation for this attempt is reviewed in detail below. Briefly though, there is a great deal of evidence to suggest that low-level phonological structure is intimately involved in the learning and recall of verbal materials. For instance, most errors in nonword repetition are phonemic substitutions which preserve the syllabic structure of the target (e.g., Treiman & Danis, 1988). A similar pattern is found in spontaneous speech errors

The authors thank Sophie Scott, Dave Glasspool, Rachel Harris, and two anonymous reviewers for their helpful comments on previous drafts of this report. We are also grateful to Rebecca Treiman and Sue Gathercole for supplying us with materials and supplementary information on their studies on nonword serial recall and repetition. The authors thank Prahlad Crupta who alerted them to an error in the original implementation of the model. This work was supported in part by SERC studentship 91313677 to Tom Hartley and JCI project grant SPG/9200174 to George Houghton. Correspondence and reprint requests should be addressed to Tom Hartley. Department of Psychology, University College London, Gower St., London WC1 6BT, UK.

0749-596X/96 $18.00

(Dell, 1986). However, existing models of speech production, and models of verbal short-term memory have hitherto remained separate from each other. The current work is the first substantive attempt to integrate these areas, and to explore the consequences of doing so.

EMPIRICAL BACKGROUND

The simplest way to test short-term memory for novel stimuli is by nonword repetition. Because the task is fairly easy for normal adults, its use has been largely confined to the testing of developmental and clinical subject groups. These studies show that in normal children, accuracy improves with age. Four year-olds can already repeat 70% of two-syllable nonwords correctly, and by 8 years, repetition is 90% correct. The more syllables a nonword has, the less likely it is to be recalled correctly. The majority of errors are single phoneme substitutions (e.g., "glistering" → "gristering," Gathercole, Willis, Baddeley, & Emslie, in press), and there is evidence that they are constrained by linguistic principles. For instance, Caramazza, Miceli, and Villa (1986) report that in their patient IGR's substitutions, phonemes were much more likely to substitute for others sharing the same manner of articulation. Bisiacchi, Cipolotti, and Denes (1989) reported that their patient RR's errors were similarly constrained.

A similar task is the serial recall of lists of nonwords. Even short lists of nonwords are challenging enough to test normal adult speakers. Hulme, Maughan, and Brown (1991) compared memory span for words and nonwords and found that while span varied linearly with speech rate for both types of stimulus, lists of familiar words were recalled better than lists of nonwords. The slopes for the two linear functions did not differ significantly, while the intercepts did. For lists of one syllable words, for example, span was about 5.0, whereas for one-syllable nonwords span was around 3.5. It was suggested that both word and nonword recall was supported by a short-term store employing an articulatory control process which was responsible for the consistent effect of speech rate on span. The difference between the intercepts was attributed to the availability (independent of stimulus duration) of long-term phonological representations of the familiar items.

Other data (Ellis, 1980) make it clear that, when unfamiliar stimuli are used, errors are qualitatively different from those reported in the serial recall of familiar items. Although entire syllables are occasionally recalled in the wrong serial position, more often errors involve phonemes from different target syllables recombining to form new syllables. These recombination errors have been found to be highly constrained, so that, for instance, initial phonemes tend to substitute for other initial phonemes, and phonemes are likely to share articulatory features with those they replace. In addition, consonants are more vulnerable to substitutions than are vowels.

The most detailed analyses of error types in nonword recall were reported by Treiman and Danis (1988), and these data are discussed more fully later ("Simulations"). They show a remarkable tendency for the syllabic structure of target items to be retained, even when the phonological content is jumbled. For lists of six CVC syllables, only 42% of responses (averaged over all serial positions) were correct, but 96% of errors had the same consonant–vowel structure (CVC) as the target syllables. This shows that phonemic substitutions are the most common error type, and that insertions and deletions are comparatively infrequent. Most errors were recombinations of phonemes from different list items, though there were also many extra-list substitutions. A similar pattern of results was observed for syllables with different shapes (CVC and VCC). In each case, the great majority preserved the CV-structure of the target syllables. It is clear from such findings that consonants and vowels rarely, if ever, substitute for one another.

Such recombination errors are unusual in the recall of familiar words, although they have been observed in children (Brady, Shankweiler, & Mann, 1983). By contrast, errors in which whole items are misordered occur in both word and nonword recall. A simple explanation for these findings is that the recombination errors reflect the failure of a subsystem dedicated to the retention of phonological forms. When the stimuli are familiar, this subsystem is somewhat redundant, as the forms of the targets are already well-learned (at least in normal adult speakers). Recall can thus be supported by long-term phonological knowledge, as demonstrated by Hulme et al. (1991). When access to long-term phonological knowledge is disrupted, short-term memory for the once-familiar items suffers. Patterson, Graham, and Hodges (1994) describe the phonological errors made by patients suffering from semantic dementia, a disorder which results in the progressive loss of semantic knowledge. Patients also exhibit a decline in vocabulary, presumably because semantic access to the lexicon is disrupted. The result is that once-familiar words are lost. Patterson et al. (1994) presented the patients (PP, FM and JL) with short lists of *known* (that is retained) and *unknown* (irretrievable) words which they had to repeat. The patients made a large number of phonological errors in recalling sequences of *unknown* words. The pattern of errors was strikingly similar to that reported by Treiman and Danis (1988). Patterson et al. conclude that the words' long-term phonological representations are normally supported by the semantic knowledge which has been disrupted in these patients. The data suggest that, when these *long-term* representations are no longer available, recall is entirely dependent on the same short-term store which normal subjects use to remember nonwords.

Other accounts (e.g., Caramazza et al., 1986) proposed that, far from interacting as the above data suggest, long- and short-term stores have separate phonological output systems. These accounts are motivated by case studies of patients showing particular problems with nonwords (compared to familiar words) in a range of tasks assumed to involve working memory. Both acquired (Caramazza et al., IGR, 1986; Bisiacchi et al., RR, 1989) and developmental cases (Campbell & Butterworth, RE, 1985; Hulme & Snowling, JM, 1992) have been described. In such patients, performance on nonword repetition and other tasks entailing nonword processing is more disrupted than that for words. Importantly though, measures of span for familiar items generally show some deficit.

We feel that the postulation of specialized systems for the output of familiar forms is intuitively unappealing. Why should the retrieval of novel forms require the use of different processes than those used for the production of familiar words? It is also at odds with data showing that phonological errors in nonword recall are qualitatively similar to those in spontaneous speech production (Ellis, 1980). Finally, it cannot explain why difficulties in nonword processing are so often associated with a general short-term memory deficit. A more parsimonious account would explain these problems with unfamiliar stimuli in terms of the general short-term memory deficit and the processing demands of the different tasks (i.e., nonword recall requires retention of phonological form, whereas word recall does not).

THEORETICAL BACKGROUND

One of our main aims in developing the model is that it should dovetail well with existing ones applying to related processes, such that it provides insight into how the subsystems interact, as well as the way they operate in isolation. There are two of areas of existing theory with which the new model will make contact: models of verbal short-term memory and models of spontaneous speech production. Here we review each of these areas in turn.

1. Models of short term memory. The articulatory loop model of verbal memory (Baddeley & Hitch, 1974) provides an account of a variety of well-known phenomena, such as the reduction of span under articulatory suppression, which can be understood. However, until recently, the model was limited by its lack of computational specification, especially regarding the basic issue of the representation of serial order. This problem has been addressed by Burgess and Hitch (1992) who put forward a connectionist model of the articulatory loop, using a "Competitive Queuing" mechanism (Houghton, 1990) to maintain order information. In this class of model, during learning, items in a target list become associated with the states of an internally generated dynamic context, which provides a "distributed" representation of the serial positions of the items. Recovery of the context during recall leads to the parallel activation of items which occurred in close proximity during learning. The sooner an item is to be output the more strongly it is activated, and all activated responses compete for control of output at a “competitive filter.” This filter picks out the most active response and then suppresses it, allowing generation of the next response (c.f., Estes, 1972; Rumelhart & Norman, 1982). Competitive queuing models separate out response preparation (which can occur in parallel) and response selection (which occurs serially). They are thus consonant with Lashley’s (1951) seminal observation that responses are active (at some level) before being actually generated (Houghton & Hartley, in press).

In the model developed here, we taken the Burgess and Hitch (henceforth B&H) model as our starting point with regard to the general problem of the representation of serial order in verbal short-term memory. The model has the crucial property of being able to learn and recall an ordered list of items in a single trial, and exhibits such human-like properties as a bowed serial recall curve (Burgess, 1995), and a proneness to ordering errors such as paired transpositions. However, the B&H model is limited in that (a) it can only handle familiar words, and (b) it is only concerned with the ordering of the words themselves; no mechanism is provided for the ordering of phonemes within a word.

2. Models of spontaneous speech production. One of our aims is to account for the pattern of constraints which govern which phonological errors can occur in the recall of nonwords. In models of spontaneous speech production, similar constraints are typically realized by separating phonological structure from content. Although there are many (often informal) models with this character, Dell (1986, 1988) describes a spreading activation model of sentence production which is particularly relevant to the present investigation. In this model, priming effects are responsible for phonological order errors in spontaneous speech. Forthcoming words in the intended utterance receive some activation in advance of their articulation (as in the competitive queuing models described above). This activation spreads through the network to the phonological output layer, occasionally leading to phonological errors in the output. In particular, Dell’s model shows anticipatory and perseveratory substitutions of single phonemes, which are the most frequent categories of spontaneous speech error.

In Dell’s (1986) model, the order in which phonemes are selected for output is constrained using a syllable schema. In addition, phonemes are represented by different nodes depending on their syllable position—an onset /k/, for instance, is represented by a different node from a coda /k/. To produce a syllable, the syllable schema selects the most active onset, peak and coda nodes in turn. A later version of the model (Dell, 1988) employed a slightly different mechanism using a number of wordshape (CV-structure) representations. In each case, syllable structure is represented separately from phonemic content, and the two representations interact in the production process. Dell (1988) suggests how this interaction might be implemented: “Selection could be achieved by any of several mechanisms. For example, the phoneme category nodes could be activated in series . . . each would send an increasing amount of activation to all of the phoneme nodes until one of them, the one with highest activation level to begin with, reaches some selection threshold.” Dell’s model thus postulates the existence of competitive processes at the output level, based on relative activation level, just as in the competitive queuing models discussed above. However, in order to account for observed error patterns, response competition at any syllabic position is restricted to those phonemes which can occupy that position (according to a syllabic template).

Given the finding that errors in the serial recall of nonwords are constrained in similar ways to phonemic errors in spontaneous speech, it is natural to propose that the underlying articulatory control processes should be the same in both cases. The greater incidence of errors in the nonword case may be accountable for by postulating that, having had to be learned in a single trial, their representations are more “fragile” than those of familiar, well-learned items (a similar proposal is made to account for differences in word and nonword spelling in Houghton, Glasspool, & Shallice, 1994). The use of activation-based competition to explain serial order errors makes Dell's approach to speech production the obvious candidate for integration with the Burgess and Hitch model. In the following section we describe a network model which achieves such integration.

DESCRIPTION OF THE MODEL

In order to incorporate phonological structure into verbal short-term memory, the model must simultaneously represent stimuli at (at least) two levels: the syllable level and the phoneme level. At the syllable level, an input stream of phonemes (representing a to-be-learned list of nonwords) is parsed into syllabic chunks. The syllable level has the task of remembering which syllable occurred at which position in the list. At the phoneme level, the identity and order of phonemes within each syllable must be remembered, so that each syllable is not only recalled at the correct position in the list, but also has the correct form. In principle, errors in recall could occur at either level. Consider, for instance, a target list, /ɡer, vɑŋ, kus, dæl, jɔb, ʃim/. Errors might involve the reordering of entire syllables, e.g., /ɡer, vɑŋ, dæl, kus, jɔb, ʃim/. Phoneme-level errors could also occur, recombining phonemes from different syllables, or introducing phonemes not present in the target syllables e.g., /ɡer, vɑs, dul, kæm, jɔp, ʃim/.

The system responsible for the single-trial learning and recall of syllable order is assumed to operate on the principles developed by Burgess and Hitch (1992) in their network model of the articulatory loop. This is supplemented by a lower-level component representing syllabic structure and content, which tracks the input stream in real time, parsing it into syllables, and generating a representation capable of supporting the immediate repetition of novel stimuli.

The hierarchical nature of the model, makes its dynamics somewhat complex. To avoid overloading the reader with detail, we therefore provide a largely informal description of the model in the main text, aimed at establishing a clear intuitive picture of how it operates. A full mathematical description of the model is provided in Appendix 1.

Architecture

The model is implemented as a neural network, in line with relevant previous work (Burgess & Hitch, 1992; Dell, 1986; Houghton, 1990). The architecture of the model is shown in Fig. 1. Figure 2 is a more detailed diagram showing the network's nodes and connections (many of which have been omitted for clarity). The model contains a set of "uncommitted" units for the encoding of syllables as they appear in the input. The phonological form of these syllables is encoded in two separate pathways (see Fig. 1):

1. The *content pathway,* directly links the syllable unit to units representing its constituent phonemes.
2. In the *structural pathway,* connections are via a general syllable template. Excitatory connections of fixed strength link nodes representing 'slots' in the syllabic template (discussed in detail below) to the nodes representing phonemes that can fill those slots.

The other connections in the model have variable weights, which are set during learning (described below). We follow Burgess and Hitch (1992) in postulating that these connections used for short-term storage are temporary. Their strengths decay over time to prevent saturation of the system so that the same weights can be used again and again.

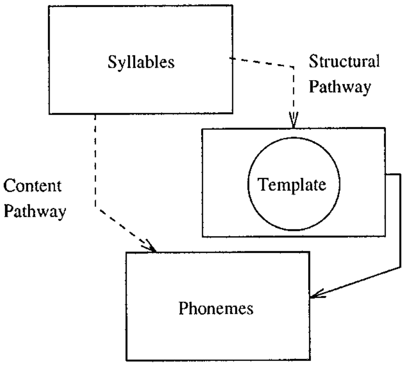

FIG. 1. An outline of the model's structure.

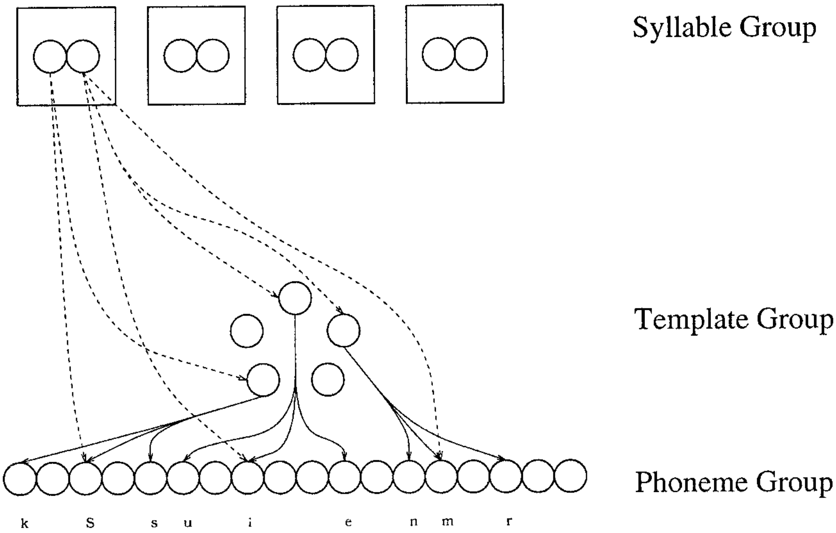

FIG. 2. A more detailed diagram of the model's structure showing nodes and connections.The dashed lines represent temporary weights, the solid lines permanent connections.

The Phoneme Group

The simplest group are the phoneme nodes which represent phonemes perceived during learning, and articulated during recall. In the current implementation there are 47 nodes representing 20 vowels, 24 consonants and the clusters /sp/, /st/, and /sk/ which are treated as single consonant phonemes (following Fudge, 1968, see The Syllable Template below).

The Syllable Group

The syllable group is comprised of pairs of nodes, each pair able to represent a single syllable. One node of each pair is associated with the onset of the syllable, the other with the rhyme (see Fig. 3). At any time, only one node of each pair has all of the activation associated with that syllable. The division of syllables into onset and rhyme is supported by a variety of linguistic (e.g., Fudge, 1969) and psycholinguistic data, including word games (Treiman, 1986) and developmental data (Goswami & Bryant, 1990), and is likely to be a feature of any plausible speech model. In the present context, this division is theoretically necessary because most of the consonants can occur in both pre- and post-vocalic positions. If a single node were used to represent a syllable, a sequence like /pæt/ would be represented identically to /tæp/, since both contain the same phonemes and involve the same "slots" in the syllable template (discussed below). Thus the onset/rhyme division is necessary to distinguish such syllables from each other;[1] it also has empirical consequences in terms of observed error types. Without it, the model would show large numbers of transpositions from pre-vocalic to post-vocalic positions, which is at odds with data showing the tendency of phonemes to maintain their syllable position in transpositions (Ellis, 1980).

The Syllable Template

The syllable template is intended to approximate the structure of a generalized (putatively universal) "syllabic gesture," based around the notion of "sonority." Sonority is a linguistic

---

[1] An alternative to our scheme, one used by Dell (1986, 1988), is to represent the "same" pre- and post-vocalic consonant separately in the phoneme group, so that /pæt/ and /tæp/ would not actually be the reverse of each other (at the phonemic level). As Dell acknowledges, this position-specific code is an unsatisfactory solution, since it fails to capture any relationship between pre- and post-vocalic realizations of the same phoneme. Following Houghton (1990), we prefer to avoid reliance on position specific coding.

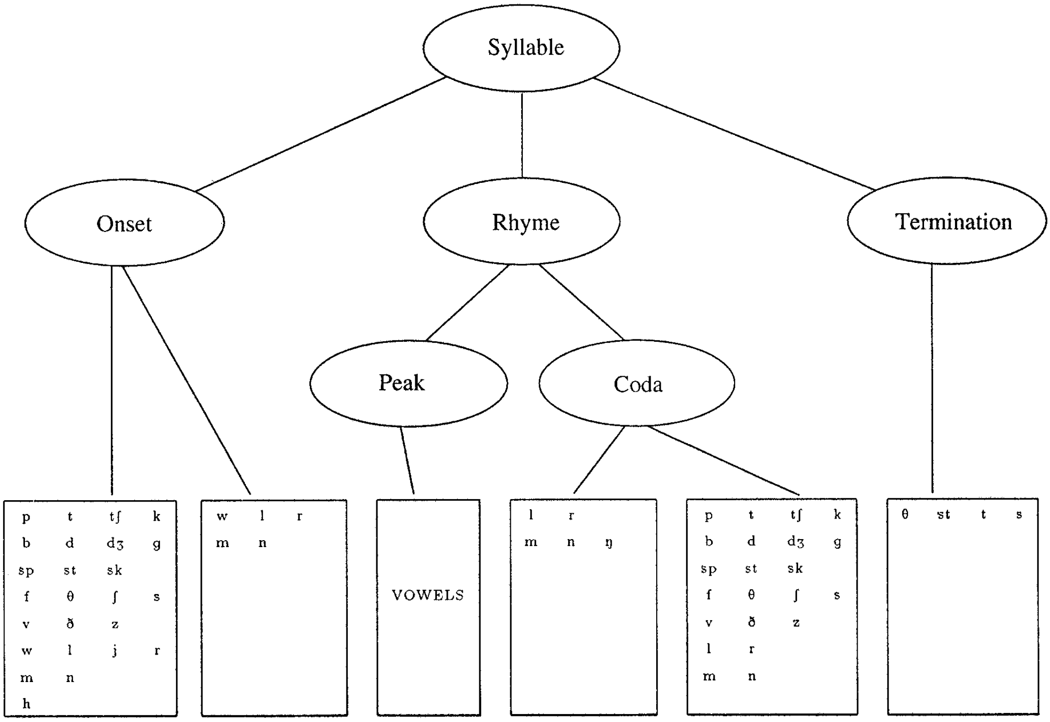

FIG. 3. English syllable structure as described by Fudge (1969).

quality associated with the phoneme, which is related to the energy of the sound it represents. A sonority principle imposes constraints on the way in which phonemes can be ordered in a syllable (see e.g., Selkirk, 1984). Phonemes may be assigned relative sonority values, and, in a well formed syllable, successive elements increase in sonority toward a peak (the vowel), and then decrease again. Obstruents have low sonority, nasals and liquids have intermediate values, and vowels have highest sonority. The sequence /flʌnt/ conforms to the principle whereas sequences like /lfʌnt/ and flʌtn/ do not. The general constraints imposed by sonority are usually supplemented by language-specific constraints. For example in English, the cluster /tl/ cannot occur in the onset part of the syllable (Fudge, 1969). It is the more general constraints on phonological structure imposed by the sonority principle which will be represented in the current model.

The detailed form of the template implemented in the current model is based on work by Fudge (1969). Fudge proposed that English syllables conform to the structure shown in Fig. 3. The syllable has onset, rhyme, and termination constituents. Each of these subsyllabic constituents is comprised of a number of slots, each of which can be filled by any one of a subset of phonemes, or remain empty. Note that the clusters /sp/, /st/, and /sk/ can occupy a single slot, the remaining basic elements of the grammar are phonemes.

The five slots in Fudge's onset and rhyme are implemented in the current model, shown in Fig. 4. The termination position is reserved mainly for inflectional endings and is omitted from the current implementation. As shown in Fig. 2, each slot in the template is represented in the network by a node linked by permanent top-down connections to a subset of phoneme nodes. We will assume here that these connections represent part of a speaker's long-term phonological knowledge. The particular phonemes associated with each slot are listed next to it in Fig. 4.

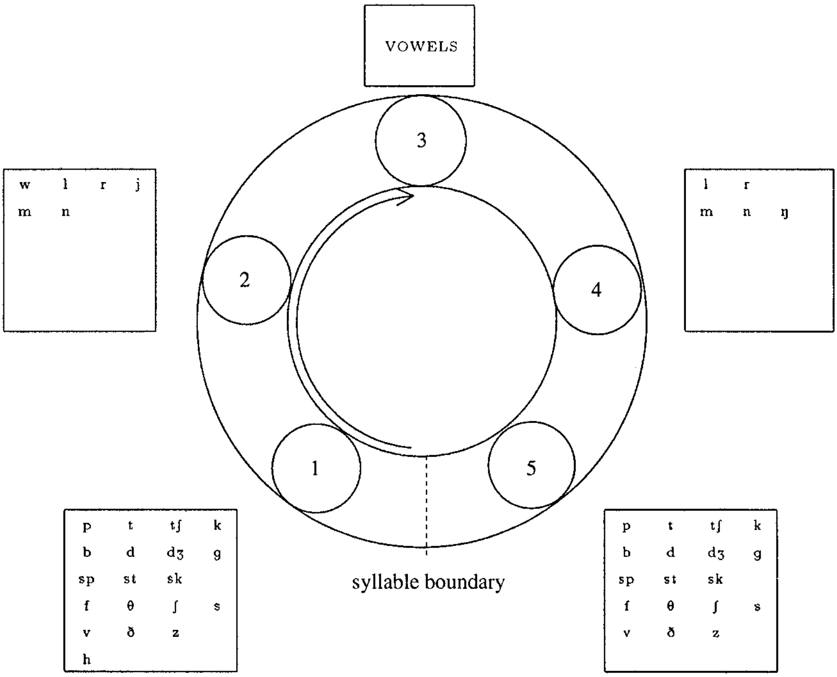

FIG. 4. The structure cyclical syllabic template used in the current model.

The relationship between phonemes and slots has been simplified, so that each consonant is associated with no more than one pre- and one post-vocalic slot. The sonority of segments associated with successive slots increases until slot 3, after which it declines. Because few legal syllables violate the sonority constraint, most can be described in a single cycle by progressing in clockwise direction to the next matching slot for each phoneme encountered in sequence. The legal nonsense syllable /flʌnt/, for example, uses slots, 1,2,3,4, and 5; /pæt/ uses 1,3, and 5. At the same time, a large number of potential sequences that are not legal syllables (e.g., /lfʌnt/) cannot be described in this way. In general, sequences violating the sonority principle would require more than one cycle. A few legal (English) syllables cannot be "parsed" in a single cycle, for example /ɪts/. Such syllables either involve the use of Fudge's termination position, or occasionally contain successive phonemes from the same slot. Because language-specific constraints are not implemented, a number of phonotactically illegal (in English) syllables can also be parsed, for example /tlɔrk/. The syllable template is clearly something of a simplification, and it fails to capture all the phonological constraints on English. It is, however, sufficient to capture the general constraints arising from the sonority principle and to represent most legal syllables.

The template can be also used to parse a continuous stream of phonemes into syllables. Whenever a new cycle begins, a new syllable has been encountered. Table 1 shows how a continuous sequence ("Rumelhart and Norman") is parsed into syllable-sized chunks.

Learning and Recall

Simulations consist of two phases: learning and recall. During learning, the model is given one presentation of a sequence of phonemes constituting a set of nonsense syllables. The model is then required to recall the input sequence as best it can. The presented and recalled sequences are then compared and any errors the model makes counted and classified.

Learning

During learning, the network is presented with streams of phonemes which it parses into

TABLE 1

A CONTINUOUS STREAM OF PHONEMES IS PARSED INTO SYLLABLE-SIZED CHUNKS BY ASSIGNING EACH PHONEME TO THE NEXT AVAILABLE CLOCKWISE SLOT

Sequence	Slots used
/rʊm/	2,3,4
/el/	3,4
/hɑrt/	1,3,4,5
/ænd/	3,4,5
/nɔr/	2,3,4
/mən/	2,3,4

syllables using the syllable template discussed above. As each phoneme is presented, a single node in each group is activated; that is, the phoneme node representing the input phoneme, the template node representing the next clockwise matching slot in the syllabic template, and one syllable node. We describe how each group is updated in turn.

Activation of phoneme nodes. At each time step, a single active node in the phoneme group represents the current input phoneme. All the other phoneme nodes are inactive.

Activation of syllable nodes. Uncommitted syllable nodes are activated in turn. Once an onset/rhyme pair have been used to encode a syllable, they are not used again. When the active template node (see below) represents slots 1 or 2, the onset node carries all of the activation of the syllable unit, otherwise the rhyme node is activated. The active syllable unit is changed each time a cycle of the template is completed. Using the weight change rule described below, each onset/rhyme pair encodes the structure and content of a single input syllable.

Activation of the template nodes. Template nodes are activated bottom-up by phoneme nodes. In principle, each phoneme node can activate any slot at which it can legally occur. However, most consonants can occur in either pre- or post-vocalic positions. We therefore assume that which of these is activated depends on the preceding phonological context, as represented by the previously activated template node. The template node activated by any input phoneme is the next available clockwise slot which can be filled by that phoneme. Consonants following the vowel will therefore activate post-vocalic consonant slots, if possible. In Appendix 2, we show how the model can be trained to behave in the manner described, by exposure to a corpus of legal English syllables.

Figure 5 shows how various nodes in the network become active in response to the presentation of a nonword. As this occurs, the weights on the connections between them are altered according to a Hebbian learning rule, such that concurrently active nodes have their (temporary) connections strengthened. In addition, the strengths of the temporary connections also passively decay toward zero. For any two nodes (designated u_i and u_j) having activation values of α_i and α_j respectively, and linked by a temporary connection of weight w_{ij}, the change in the weight at each timestep (Δw_{ij}) is given by:

$$\Delta w_{ij} = \alpha a_i a_j - (1 - \delta) w_{ij}, \qquad (1)$$

where δ is a decay term ($0 < \delta < 1$), and α is the learning rate.

In some of the simulations described below,

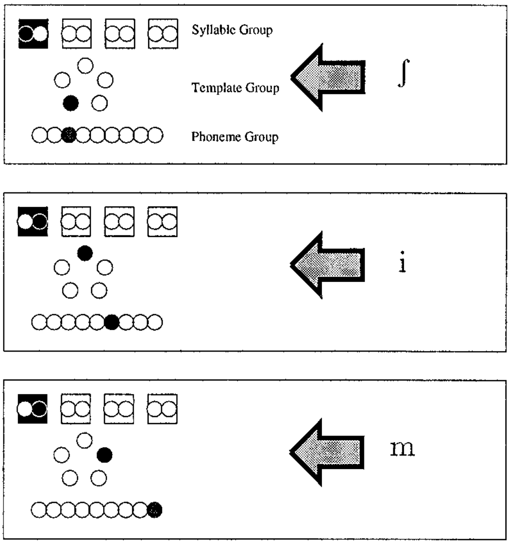

FIG. 5. During learning, the changing state of the input stream (represented by phoneme nodes) determines the activation state of nodes in the syllable and template groups. This diagram shows how the state of the network changes during presentation of a monosyllabic nonword (/ʃim/). The activation of each node is represented by its shading (the darker the more active). From top to bottom successive frames show the state of the network as each phoneme appears in the input stream.

the model is run in simulated real time. It is therefore convenient to express the decay rate as a half-life h, the time taken in seconds for the weights to decay to half their initial values. For a given timestep (of duration d), δ is given by:

$$\delta = 0.5^{\frac{d}{h}}. \tag{2}$$

The result of learning is that, for each syllable in the input, excitatory connections simultaneously become established between (i) syllable units and phoneme units (encoding phonemic content) and (ii) syllable units and template units (encoding syllabic structure).

Recall

The aim of the recall process is to recreate the serial pattern of activation over all the nodes in the network that occurred during learning. The serial pattern of activation over the syllable nodes represents the recalled order of the syllables themselves. Item order errors would occur at this level. The sequence of phoneme and template node activations represents the phonological form of the recalled syllables. Errors of syllabic form and content occur at this level. No learning takes place during recall (i.e., $\alpha = 0$), and hence the temporary weights established during the learning phase decay exponentially, according to Eq. (1) above.

The way in which the various groups of nodes become activated during recall is described in turn. The overall pattern is shown in Fig. 6.

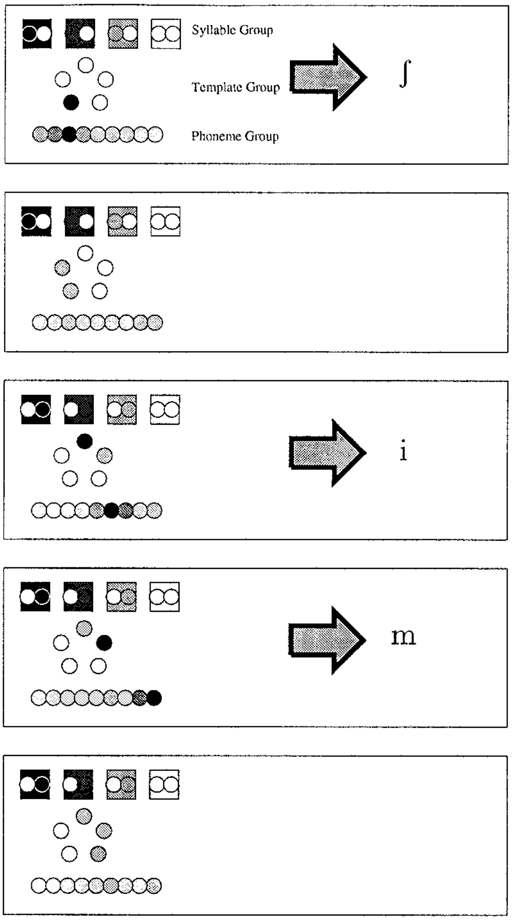

FIG. 6. Diagram illustrating activation states of the network during recall of a nonword (/ʃim/). As in Fig. 5, the degree of activation is represented by the density of the shading.

Activation of the syllable nodes. During recall, activation of the nodes in the syllable group is assumed to be controlled by the competitive queuing mechanism described by Burgess and Hitch (1992). The most important point about this mechanism is that *it causes a number of syllable nodes to be active in parallel,* with a gradient of activation over them such that syllables are more active the nearer they occurred to the current target syllable during learning. After each syllable is output, the most active node is suppressed. An additional constraint is that, as during learning, all the activation associated with an onset/rhyme node pair is either in the onset node or the rhyme node. The same member of each pair will be active in all concurrently activated onset/rhyme pairs. This two-stage retrieval process is consistent with the experimental findings of Meyer (1991) concerning the time-course of phonological retrieval in speech production.

Activation of the syllable template. Template nodes receive input from onset/rhyme nodes. During learning, connections will have been formed between these nodes and the syllabic slots that were activated by the input syllable. In order to produce a syllable at recall, the syllabic template must be accessed such that the template nodes associated with the target syllable "fire" in series. This is achieved by applying external input cyclically to the template group. This cycling input is a very simple dynamic pattern which moves serially through the template, once for each syllable to be recalled. The level of the external input (ϵ, see Appendix 1) is adjusted to offset the effects of weight decay, so that the net input to the template is the same regardless of duration of the stimulus material. It is desirable that it be sufficient to fully activate (or "fire"[2]) a template node only if that node is currently receiving input from the (strongly active) onset/rhyme pair representing the current syllable. When this condition is met, the combination of "endogenous," cyclic input, with the input from the syllable units via the learned weights, allows the syllabic structure to be recalled.

Activation of phoneme nodes. It is the serial pattern of activation of the phoneme nodes which constitutes the model's output, its attempted reconstruction of the input experienced during learning. A phoneme node receives input from both the structural and content pathways. Typically the weights developed by the network during learning will be such that neither is sufficient in isolation to activate any node beyond a threshold at which it competes to be output (see Appendix 1 for further discussion). However, a phoneme node will become active if it receives sufficient combined input from *both* structural and content pathways. Note that, due to the competitive queuing recall algorithm, syllable group nodes will be active in parallel. In addition, firing of any of the template nodes will provide input to all of the phoneme nodes to which it is connected. This leads to parallel input to the phoneme nodes. Most input will be received by phonemes which can fill a particular syllabic slot, especially those which appeared close to the current target syllable in the list.

The activation of nodes throughout the network is subject to noise. Noise becomes increasingly important in determining which of the activated phonemes is selected as the delay between learning and recall increases, because the level of input from the syllable nodes declines due to weight decay. This produces errors, which are discussed in more detail later. The performance of the model at recall is thus dependent on the amount of decay that has taken place in the strengths of the temporary connections. This is in turn dependent on the duration of the list or item to be recalled, and the rate at which it is articulated during learning and recall. For example, if recall is faster than presentation, then the weights associated with the representation of the segments near the end of the sequence will have decayed less (at the time of their recall) than those associated with the beginning. Such interactions like these, between serial position and rates of presentation and recall, while worthy of further investigation, are not the focus of attention in this study. In the simulations reported below, recall is paced at the same rate as presentation. For polysyllabic speech, the template is cycled continuously; for monosyllabic items in a list, pauses are allowed between recall of different items during which decay continues.

SIMULATIONS

In the simulations described below, the syllable sequencing mechanism is assumed to operate perfectly. The results thus show effects that are specific to the recall of unfamiliar materials. *In addition* we would also expect to observe effects of errors in the syllable sequencing mechanism, which we assume is like that postulated for the sequencing of familiar words by Burgess and Hitch 91992). The B&H model predicts effects of both list length and serial position, thus longer lists will be more prone to syllable level errors (typically transpositions), which will be concentrated toward the middle of the list. The simulations presented here show how the additional difficulty of remembering the phonological forms of novel stimuli is expected to affect recall.

Quantitative Measures of Performance

The following simulations show how the model's performance changes for lists of varying length. If a hierarchical model of sequenc-

---

[2] Unlike the other nodes in the network, the activation function of the template nodes is "all-or-none" so that they are either very active (when receiving input greater than a threshold) or inactive.

ing is to be adopted, then most errors must be at the phonemic sequencing level. Span for nonwords is much lower than span for words. If the same sequencing mechanism is responsible for the ordering of both familiar and unfamiliar items, then the difference in span must be largely due to problems in correctly recalling unfamiliar phonological forms, and not in ordering the items themselves.

Figure 7 shows the model's performance in simulations of two different experimental paradigms–nonword repetition (circles) and nonword serial recall (squares). The repetition simulations were based upon experiments by Gathercole and Baddeley (1989) and colleagues using the children's nonword repetition test (CNRep—see, e.g., Gathercole et al., in press). The stimuli presented to the network were polysyllabic nonwords from the CNRep (for example /slædiŋ/, trumpətin/, pʒplɪstərɒnk/). The serial recall simulations involved CVC monosyllables. In both cases, the length of the input sequences can be characterized in terms of the number of syllables they contain (*x*, axis). The black symbols show the proportion of entire input sequences (lists/polysyllabic nonwords) that were recalled correctly for inputs of different lengths. The white squares show the proportion of individual items (monosyllables) that were recalled correctly.

The difference between the repetition and serial recall simulations is that in repetition the nonwords were presented and recalled as continuous phonemic sequences, whereas pauses were allowed between items in the serial recall case. This results in a longer duration (and thus greater weight decay) between the learning and recall of each phoneme in serial recall compared to repetition.

The fewer syllables in a sequence contains, the more chance there is it will be recalled correctly. This is true of both lists of monosyllables and polysyllabic nonwords. In each case the length effect results largely from decay in the temporary weights in the network. The longer the list, the more decay takes place, and the more impact noise has in deciding the outcome of competition amongst the phoneme nodes. Clearly, a longer sequence also demands that more phonemes are selected, and so there is a

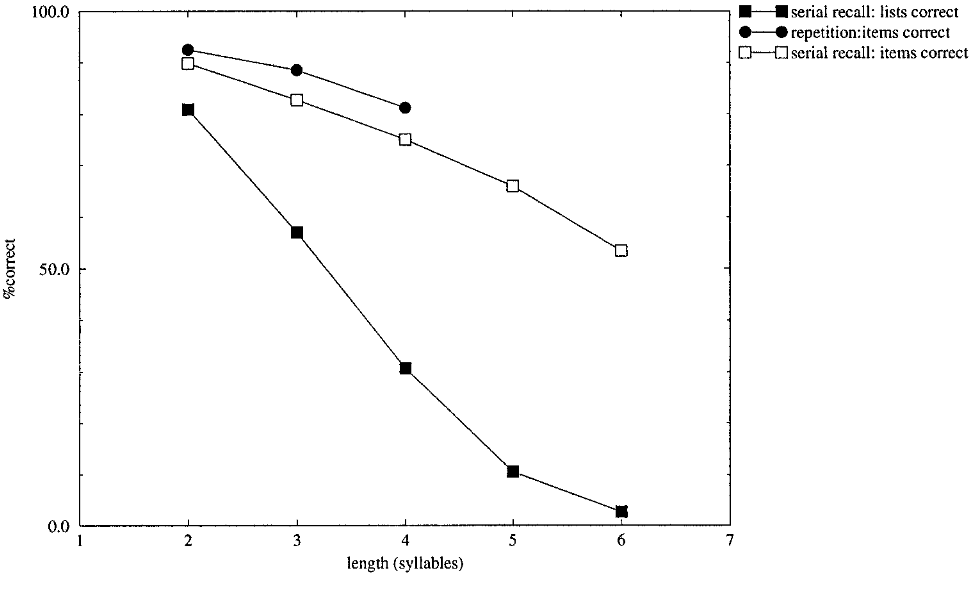

FIG. 7. Graph showing proportions of correct responses in simulation of nonword repetition for items of varying lengths in syllables (black circles). Each point is plotted from 500 presentations. The results of nonword serial recall simulations are plotted on the same axes: both individual items (white squares) and lists (black squares) were scored for lists of varying length. Each point is plotted from 1500 list presentations.

greater opportunity for an error to occur. Because the lists used in serial recall have greater duration than the continuous utterances used in repetition, performance on the serial recall task is considerably worse, even when the stimuli have the same number of syllables.

Nonword span can be determined by interpolation of the proportion of *lists* recalled correctly. This gives a value of approximately 3.2 syllables, a number that agrees quite well with the experimentally determined figure (about 3.5, Hulme et al., 1991). However, bearing in mind that we would also expect errors at the syllable level to have an impact on span, the figure is a little on the low side. It is worth noting though that Hulme et al. (1991) used a rather more conservative measure of span.

Currently, we are aware of no adult experimental data which can be compared directly with the model on the nonword repetition task. Most studies using the task have involved either children or neuropsychological patients. The model's performance is comparable to the oldest children tested and shows the characteristic length effect. Later, we will show how, by increasing the rate of weight decay, the model can simulate neuropsychological and developmental data (see Modeling a Damaged/Developing Store). We suggest, on the basis of findings relating to digit span, that such rapid decay is a characteristic of the developing or damaged short-term memory system.

Qualitative Analyses

To explore the degree to which constraints on the model's errors are similar to those operating in human short-term memory errors, in this section we analyze qualitative features of the model's output and compare them to analyses of responses obtained from human subjects.

The most detailed empirical data comes from Treiman and Danis (1988). We simulated this experiment using the same 30 lists of six monosyllabic CVC nonwords they used (Table 2 gives examples). For such lists, the overall error rate was somewhat lower in the simulation (46% of responses) than in the experimental data (57% of responses). However, the model does not produce item order errors (approximately 6% of responses in the experiment), hence the error rates are reasonably comparable. For more detailed comparison with the data, the model's responses were analyzed in three ways: first, we analyzed the CV-structure of the recalled syllables, counting the numbers of each type; second, we categorized the errors in terms of substitutions, deletions etc., establishing the proportions of each; finally, we analyzed the responses in terms of the origin of their phonemes.

TABLE 2

THREE EXAMPLES OF STIMULUS LISTS FROM TREIMAN AND DANIS (1988, EXPERIMENT 1)

ger vaŋ kus dæl jɔb ʃim
ðʌp tʃel duk dʒaɪn hur zæŋ
fim wɔr tʃɔɪd ʃuð θɪl jæŋ

CV-Structure

The CV-structure of each response was recorded. The results are shown in Table 3. In Treiman and Danis (1988, Experiment 1) study about 94% of subjects' responses were CVC syllables (96% of errors in which an attempt at recall was made). The rates at which other syllable types occurred were not reported, however it is clear that some erroneous responses which did not share the CVC structure of the targets were produced, so it is important that the model also produces such responses. Table 3 shows that the syllabic template in the model is playing its functional role effectively: the great majority of responses (87.47%) had the same CVC structure as the target items. It also shows, in line with the data, that a small proportion of responses with other structures were produced.

Substitutions, Insertions, and Deletions

Errors were categorized into substitutions, insertions, deletions and so on. These results are shown in Table 4. Empirical data from normal adult subjects is not available for comparison. There is, however, reason to be fairly confident that the model's predictions will prove to be broadly compatible with future empirical findings: similar analyses have been carried out in studies of nonword repetition in both children

TABLE 3

The Proportion of Responses Showing Various Consonant–Vowel Structures in Simulations of Serial Recall of Lists of Six Monosyllabic Nonwords

CV-Structure	% Responses
CVC	87.47
VC	1.53
CV	2.62
VCC	0.09
CCV	0.09
CCVC	2.99
CVCC	4.34
V	0.02
Other	0.85

(Gathercole et al., in press) and patients with disorders of output phonology (Bisiacchi et al., 1989). In each case, single phoneme substitutions accounted for the majority of errors, with deletions and insertions being relatively much less frequent. Given the similarity of the experimental paradigms, it would be surprising if errors in serial recall of nonwords differed greatly from this pattern.[3]

Treiman and Danis' data alone clearly show that whatever the make-up of the other error types, insertions and deletions are much less frequent than errors which do not alter the number of phonemes (i.e., single and multiple phoneme substitutions). Table 5 presents data regarding the phonemic length of the model's erroneous responses. Here, human data are available for comparison: Treiman (in press) found that errors preserving the number of phonemes in target stimuli accounted for 83.7% of adults' errors. In the model's output (73.13%) of errors had three phonemes. Occasionally though, subjects (particularly younger subjects) produced errors which changed the number of phonemes. the model also produced such errors (often simple phonemic insertions and deletions). These come about principally through interactions between adjacent syllables (see Error mechanisms). The stimulus lists used by Treiman (in press) contained fewer target syllables, than those used by Treiman and Danis (1988) and in the simulation.[4] The stimuli in the simulation thus allowed more interaction, and afforded greater opportunity for insertions and deletions, hence the higher proportion of errors of these types in the model's output than in the adult subjects' responses.

TABLE 4

Distribution of Error Types in Simulations of Serial Recall of Lists of Six Monosyllabic Nonwords

	% responses
Single phoneme errors	
Substitutions	28.64
Deletions	3.50
Insertions	2.94
Multiple phoneme errors[a]	
Substitutions	4.78
Substitution + insertion	4.61
Substitution + deletion	1.09
Deletion	0.03
Insertion	0.09
Deletion + insertion	0.09
Complex errors[b]	0.04

[a] Not including errors which retained none of the target phonemes and errors in which none of the target phonemes appears in the same position in the response.

[b] Errors which retained none of the target phonemes and also errors which could only be accounted for in terms of more than two single phoneme errors, e.g. /kus/→/stu/.

Note. The scheme used is similar to that of Bisiacchi et al. (1989).

Source of Constituent Phonemes in CVC Errors

Following Treiman and Danis (1988), we categorized each CVC error on the basis of the origin of its phonemes. Different suffixes are used for phonemes which come from different target syllables: if each of the phonemes in a response belonged to a different target syllable, it would be designated $C_aV_bC_c$. If the initial consonant came from one target syllable, but

---

[3] Of course the overall error rates in serial recall may differ substantially from those in the repetition studies; there are clear differences in both the short-term memory capacities of the subject populations used, and the duration of the material to be stored.

[4] In Treiman's (in press) developmental study, adult performance on a nonword serial recall task was compared with that of younger subjects. The lists presented to all subjects contained three target items (CVC nonwords), but in an attempt to equate overall performance, lists were tailored to the subjects' memory spans by the addition of "padding" digits for older subjects. Hence the stimulus materials used differed substantially from those used in the Treiman and Danis (1988) study.

TABLE 5
THE PROPORTION OF ERRORS WITH DIFFERENT PHONEMIC LENGTHS

Error type	% of total errors	
	Adults	Simulation
Omitted syllable	10.9	—
1. Phoneme syllable	0.0	0.07
2. Phoneme syllable	2.0	10.11
3. Phoneme syllable	83.7	73.13
4. Phoneme syllable	3.2	16.00
5. Phoneme syllable	0.1	0.68

Note. Data from Treiman (in press) are shown for comparison 3.

the remaining vowel and final consonant came from another, the response would be coded as $C_bV_aC_a$. If all of the phonemes in a CVC response originated from the *same* target item, it would be coded $C_aV_aC_a$. Phonemes not on the target list are represented by the suffix, *x*, so that a response in which the initial consonant was not in the list (and the remaining phonemes originated from the same list item) would be designated $C_xV_aC_a$. This method of categorizing errors shows how different parts of the syllable are prone to different types of error, and the rates at which various complex multiple substitutions occur. Table 6 shows the proportion of responses associated with each error type observed in the simulation, along with the proportions of correct responses, and responses not categorized (non-CVC responses). For comparison, the experimental proportions (averaged over 36 subjects) are also shown (Treiman & Danis, 1988). This pattern of errors seems to be a replicable feature of the task. Treiman (in press, Experiment 4) reported the errors from adults tested as part of a developmental study. Although the developmental comparison required methodological changes from the Treiman and Danis (1988) study, a very similar pattern of phonological errors emerged. These data are also shown in Table 6.

Almost all of the errors Treiman and Danis

TABLE 6
RESPONSES CATEGORIZED ACCORDING TO THE SCHEME ADOPTED BY TREIMAN AND DANIS (1988; T&D)

Type of response	% of all response		
	Experiment		
	T&D (1988)	Treiman (1994)	Simulation
Correct	42.64	44.17	54.18
$C_bV_aC_a$	14.68	10.59	9.13
$C_aV_aC_a$	6.16	9.58	0.48
$C_aV_aC_b$	5.32	4.27	6.94
$C_xV_aC_a$	4.74	3.96	4.67
$C_aV_bC_c$	4.43	0.80	1.42
$C_aV_bC_a$	3.49	3.33	4.62
$C_aV_xC_b$	2.87	3.75	0.54
$C_aV_xC_a$	2.15	4.83	1.58
$C_xV_aC_b$	2.04	0.80	0.89
$C_aV_bC_x$	1.85	0.69	0.30
$C_aV_aC_x$	1.53	1.15	2.27
$C_xV_xC_a$	1.10	1.18	0.14
$C_aV_xC_x$	0.51	0.90	0.07
$C_xV_aC_x$	0.48	0.24	0.21
$C_xV_xC_x$	0.26	0.38	0.02
Non-CVC error	5.77	9.38	12.54

Note. The averaged experimental data (from T&D, Experiment 1, and Treiman (in press, Experiment 4) are included for comparison. CVC errors are broken down by the putative origin of their constituent phonemes (see text), correct responses and non-CVC responses have also been included. The two experiments employed different stimuli. The simulation used the same stimulus lists as T&D.

(1988) observed were CVC syllables. Yet, the variety of errors in striking—some combine the vowel of one item with the consonants of another; others combine consonants from two different stimulus items with a vowel which was not in the list at all, and so on. It is important that a model accounts for this variety. In the current model, contextual and non-contextual substitutions arise through the operation of the same mechanism—the cyclical access of the syllabic template. It is an essential component of the model, required to order segments for output. Under conditions of noise and trace decay, however, it is capable of producing all of the observed phoneme-level error types, including single-phoneme substitutions (e.g., $C_xC_aC_a$) and those involving phonemes from multiple sources (e.g., $C_aC_bC_c$, $C_aC_bC_x$). Although the overall proportion of CVC errors is somewhat lower in the model's output than in the experimental data, the simulation results show a similar distribution of error types. This is illustrated by the scatterplot in Fig. 8a, in which for each category of response, the observed and simulated proportions from Table 6 are plotted against one another ($r^2 = 0.94$ for these data sets). Figure 8b shows the same simulation data plotted, this time, against experimental data from Treiman (1994) ($r^2 = 0.94$).

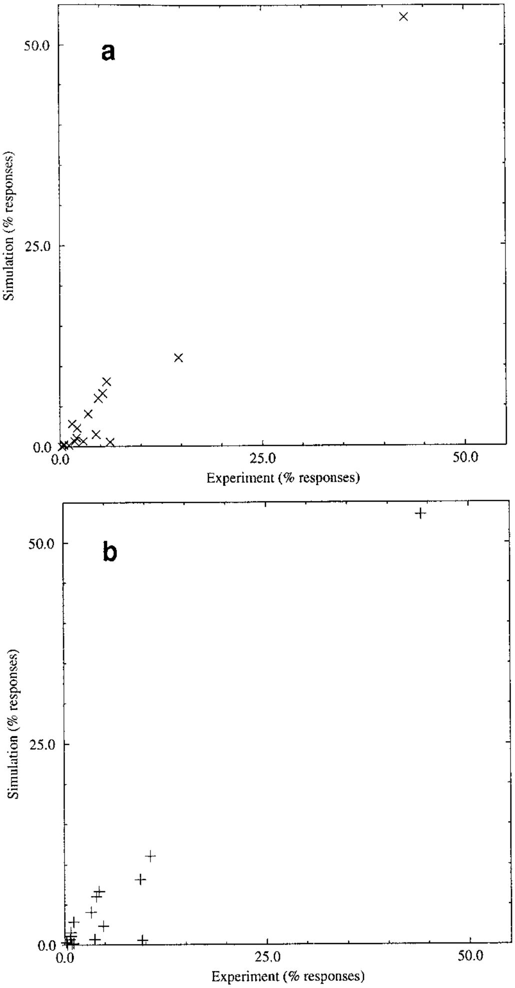

FIG. 8. Scatterplots in which, for each response categories reported in Treiman and Danis (1988) the experimentally observed rate (averaged over subjects) is plotted against the predicted rate (from the simulation output). Experimental data are from (a) Treiman and Danis (1988, $r^2 = 0.96$) and (b) Treiman (in press, Experiment 4; $r^2 = 0.94$). Points are close to the diagonal indicating that the simulation output shows a similar distribution of error types to that observed experimentally.

Treiman and Danis (1988) placed great emphasis on their finding that in the serial recall of nonwords, errors most often involve the recombination of an intact onset with an intact rhyme (for CVC syllables these are categorized as $C_bV_aC_a$ errors). Such errors are a good deal more frequent than, for example, $C_aV_aC_b$ errors, where the boundary between recombining fragments does not coincide with the onset/rhyme boundary. The model's output shows the same asymmetry ($C_bV_aC_a$ errors are more frequent than $C_aV_aC_b$), although it is somewhat less pronounced (the model produces fewer $C_bV_aC_a$ and more $C_aB_aC_b$ than the human subjects).

The difference in frequency between the two error types shown by the model's output is largely due to structural properties of the stimulus items (the same ones used by Treiman and Danis, 1988, Experiment 1): while most of the pre-vocalic consonants (all of them in some of the lists) were obstruents (associated with slot 1), there were equal numbers of nasals, liquids, and obstruents in the post-vocalic positions. Because the nasals and liquids are both associated with slot 4, they are less vulnerable to substitutions (see point 1 above). Statistically, the effect is to make post-vocalic substitutions less likely.

The account given above goes some way to explaining the different rates of recombination errors in Treiman and Danis (1988, Experiment 1). However, it applies only to the particular stimuli they used. It could easily be tested by using stimulus lists in which, for example, all of the initial consonants came from slot 1 of the syllabic template and all of the final consonants from slot 5. For lists like these, the model predicts that there would be practically no difference in the vulnerability of initial and final consonants. Therefore, we would expect this manipulation to reduce the difference in rates of $C_bV_aC_a$ and $C_aV_aC_b$ errors.

However, the simulation results do not show such a marked difference in the rates of the two error types as is observed experimentally. While the structural properties of the stimulus items did tend to produce more substitutions involving pre-vocalic segments, this cannot completely account for the difference in the observed frequencies of $C_bV_aC_a$ and $C_aV_aC_b$ errors. Thus, a modification to the current model may be necessary. The one that suggests itself is that competition occurs between onsets and between rhymes, rather than between syllabic units. This account differs in that it involves two competitive queues at the syllable level (one for onsets, one for rhymes) rather than the single queue (of syllables) postulated in the current model. Both queues would be subject to transposition errors, and so errors involving the recombination of onsets and rhymes from different syllables would be frequent. This adapted model would provide a more intuitively appealing account of the errors which most interested Treiman and Danis (1988). The current model, however, in suggesting only one sequencing mechanism at the syllable level, is simpler than one entailing the independent sequencing of onsets and rhymes.

For completeness, we have included the $C_aV_aC_a$ errors in our analysis. Though the simulation did yield a small number of $C_aV_aC_a$ errors, these all involved rearrangements of phonemes from the same list item.[5] By contrast, most, if not all of the experimentally observed $C_aV_aC_a$s, were item-order errors. Such errors will result from the breakdown of the higher level item ordering processes which, in the current implementation, are assumed to function perfectly. Thus, it is not surprising that $C_aV_aC_a$ errors are considerably underrepresented in the model's output (there are also correspondingly more correct responses).

The simulation showed that the detailed pattern of response types made by the model is similar to that observed in normal adult subjects. In the next section we explore the ways in which the model predicts nonword recall will be affected by a reduction in short-term memory capacity.

Modeling a Damaged/Developing Store

There are a number of ways in which the overall performance of the model can be disrupted, for instance by increasing the level of noise in the system. Alternatively, damage to the short-term phonological store may be modeled by reducing the half-life of the temporary weights in the network. The latter approach is explored below, in an attempt to provide an account of the often observed association between subnormal span and specific deficits of nonword processing. Such observations indicate that where the phonemic structure and content of an unfamiliar stimulus is particularly vulnerable to be disrupted in short-term memory, memory for the serial order of familiar items is also poor relative to the normal adult population. It seems reasonable, therefore, to interpret the particular problems that some subjects experience with nonwords as a reflection of a more general difficulty in maintaining a stable trace over time.

In the current model, trace decay results from the exponential decline in the strengths of temporary connections throughout learning and recall. Temporary weights are involved in the rep-

---

[5] Typically a pre-vocalic consonant is repeated in the post-vocalic position (e.g., /vʒf/ → /vʒv/, /tʃɔɪd/ → /tʃɔɪtʃ/). Errors like these arise in the same way as noncontextual substitutions. It is only coincidence that the substituting phoneme happens to be one from the same syllable. Treiman (personal communication) reports that errors like these accounted for few, if any of the experimentally observed responses so categorized.

resentation of serial order at the level of the items themselves (see Burgess & Hitch, 1992), as well as their phonemic structure and content (but, crucially, *not* in the representation of the phonological form of familiar words—see Extending the Model: Short-Term Memory and the Lexicon). An increase in the rate of weight decay could explain the increased incidence (relative to normal adults) of errors at both levels observed in patients with acquired deficits of output phonology, with some developmental syndromes, and in children. The addition of pathologically rapid decay of activation in Dell's model of sentence production (Dell, 1986) has been successfully used to provide an account of paraphasias in deep dysphasia (Martin, Saffran, Dell, & Schwartz, 1994). In the current model, information about the form of the intended utterance is represented not by activations, but by connection strengths (fixed in Dell's model). However, the same principle can be used to explain output deficits in both spontaneous production and repetition: when "planning" information is lost more rapidly through decay, speech errors occur more frequently, because the competitive selection of the correct segments for output is increasingly affected by noise.

The responses made by normal subjects presented with lists of nonsense syllables typically contain large numbers of single phoneme substitutions. The same type of errors predominate when patients with acquired deficits in the short-term retention of nonwords are asked to repeat shorter polysyllabic nonwords. In the model, these errors result from increased competition among phonemes, which in turn results primarily from weight decay. We propose that such errors, in which the structural characteristics of target items are maintained while specific content is lost, are characteristic of a short-term store operating close to the limits of its capacity. It is this capacity which is reduced in the cases of both acquired and developmental difficulties with nonsense materials. Empirical data from such case studies can be compared with the output of the model when its capacity is reduced by altering the decay parameter, to determine whether this hypothesis is consistent with the available evidence.

Table 7 shows the distribution of errors in simulations of nonword repetition, in which the half-life of the temporary weights has been reduced from 5.0 s to 3.5 s. The same materials and procedure were used as in the simulations described under Nonword Repetition. The results shown are the average rates of each error (across serial positions) for lists of length 2–4 syllables. For comparison, the distribution of repetition errors shown by normal 5-year old children (the only age group whose responses have so far been analyzed in this way) (Gathercole et al, in press), patient IGR (Caramazza et al., 1986), and patient RR (Bisiacchi et al., 1989) are presented. It is clear that in each of the experimental studies and in the simulation, single phoneme substitutions predominate. The spread of error types in the data from 5-year-old children probably results in part from their linguistic immaturity—the tendency of children to reduce complex consonant clusters, for example, is well documented. The implementation of structural constraints in the model assumes that phonology is well learned.

Because memory span increases developmentally (see e.g., Chi, 1976), the difference between the performance of subjects at different developmental stages may also be accounted for in terms of an increase in decay rate. Prior to the adjustment in the decay parameter, the model's performance was comparable to the 8 year olds in the experimental study. After it, it is close to that of 7 year-old children (see Table 8).

Causes of Errors in the Model

In this section we discuss how the various error types exhibited by the model arise. Three general factors are involved in constraining errors: long-term phonological knowledge (as represented by the form of the syllabic template), the target syllable's structure (the slots it uses in the template), and the structures of other syllables close by in the target sequence. Below we summarize the basic effects of each in turn.

1. Long-term phonological knowledge. Different slots in the template are associated with different numbers of phonemes. The more phonemes available at a given slot, the greater the number of competitors at output, and the greater the chance that the wrong one will win the competition.

2. The target syllable's structure. If a particular slot is used in a target syllable, then the phoneme at that position is open to substitution of deletion. If it is not then an insertion may occur at that position. Occasionally, the co-occurrence of an insertion and a deletion at adjacent positions may make it appear that a phoneme from one slot has replaced one from another (e.g., /pɔm/ → /lɔm/). However, such coincidences are rare.

3. The structure of other syllables in the target utterance. Syllables close to the target syllable in a list influence recall by providing priming input to the syllabic template. For instance, if syllables adjacent to the target share the same structure, the priming serves to reinforce the target structure, thereby reducing structural errors. Conversely, if syllables near to the target are structurally very different, then the priming may interfere with the correct production of the current syllable.

The error types affected by these factors can be divided into substitutions, deletions, and insertions.

Substitution Errors

Table 4 shows that the majority of the model's errors were single phoneme substitutions. These errors may be subdivided into two types: contextual substitutions, where the replacing phoneme is present elsewhere in the target list, and non-contextual substitutions, where it is not. The two error types can occur together if multiple phonemes are substituted. Some errors classified as contextual substitutions may actually be non-contextual; i.e., although the substituting phoneme occurs in the list, this is not the cause of the error. It follows that, as more phonemes are used in the target sequence, there is less opportunity for an error to be classified as non-contextual. This factor will affect the relative frequencies of contextual and non-contextual errors for different lists.

The model's non-contextual errors occur because an activated template node activates all of the phonemes associated with it, as well as the target phoneme, leading to response competition. Because the target phoneme also receives input from the content pathway (whereas phonemes from outside the list receive none), it will generally win this competition. However, if much time elapses between the presentation and recall of a particular phoneme, weight decay in the content pathway will reduce its advantage in the competition. Noise will then begin to play an important role in phoneme selection and occasionally the winning phoneme will be one from outside the target list.

Contextual substitutions occur because syllable nodes are activated in parallel, and pho-

TABLE 7

THE DISTRIBUTION OF ERROR TYPES AS A PROPORTION OF ALL RESPONSES, IN SIMULATION OF NONWORD REPETITION

	% of all responses			
	Simulation	Five-year-olds	RR	IGR
Single phoneme errors				
Deletions	5.47	7	1.54	0.47
Insertions	0.20	1	0.00	1.64
Substitutions	14.20	11	9.23	11.97
Multiple phoneme errors				
Substitutions	1.47	5	0.00	3.52
Substitution + insertions	2.33	3	0.00	0.47
Substitution + deletions	2.20	9	0.77	0.23
Exchanges	0.07	0	0.77	0.70
Complex errors	0.13	—	—	—

Note. For comparison, the repetiton errors made by 4-year-old children (Gathercole et al., in press), and patients RR (Bisiacchi et al., 1989) and IGR (Caramazza et al., 1986) are shown.

nemes from the list thereby receive some input via the content pathway in advance of the point when they are to be output. If they share the same 'slot' as the target phoneme, they will have an advantage (over phonemes from outside the list) in the competition to control the output system. The closer they are to the current target, the greater that advantage. As noted, some apparently contextual substitutions are due to the non-contextual error source. In perseveratory substitutions, for example, the syllable which is the putative source of the intruding segment has already been output and suppressed, so that the only activation the phoneme node receives is via the structural pathway. Similarly, phonemes which appear in the rhyme constituent of one syllable do not receive input from the content pathway during production of an onset. Substitutions involving the apparent movement of segments from onset to rhyme or vice versa are not due to any contextual effect. Such errors are no more likely than any noncontextual substitution.

Ellis (1980), Caramazza et al. (1986), and Bisiacchi et al. (1989) have all noted that substitution errors appear to be constrained by phonological principles. Ellis (1980) showed that substituting phonemes tended to share features with those those they replaced in nonword serial recall. Caramazza et al. (1986) and Bisiacchi et al. (1989) found that their subjects' substitutions tended to share manner of articulation, a feature correlated with sonority. In the model, sonority relations are captured in the organization of the syllabic template. All phonemes with the same manner of articulation are associated with the same slot(s) in the template. Thus, when one phoneme substitutes for another in the same slot, there is a good chance that the target and intrusion share this feature value.

Deletions

Deletions generally occur in the model if a template node does not fire when the syllable in which it appeared during learning is being produced. This is more likely to occur if the input to the node is low, as it is when the slot is used only in the target syllable. For this reason vowel deletions are very unlikely, since the vowel slot is used in every syllable. Consonant deletions, on the other hand, are quite likely to occur, particularly if the immediately forthcoming items do not utilize the same slots as the target, for example:

/ʃim/, /wɒd/ → /ʃi/ ... .

In order to prevent such deletions from occurring too frequently, the "nonspecific" input to the template nodes, ϵ, must be increased for longer lists, so that the slot nodes can still be accessed (see Appendix 1). As the weights from syllable to template nodes decay towards zero, the slot nodes will only fire if ϵ is close to its maximum, at which point the firing of the template nodes is principally determined by the noise on the input lines. By this stage, the weights in the content-pathway will also have decayed to very low values, the result being that responses will bear little or no resemblance to the targets. So great will be the chance of error, that one might expect that, rather than attempt to respond with what will certainly be gobble-

TABLE 8

TABLE SHOWING THE PROPORTION OF CORRECT RESPONSES IN SIMULATIONS OF NONWORD REPETITION FOR STIMULI OF DIFFERENT LENGTHS

	% correct						
	Experiments					Simulations	
Number of syllables	4 years	5 years	6 years	7 years	8 years	halflife = 3.5	halflife = 5.0
2	70.49	75.95	76.56	87.52	90.70	87.60	90.60
3	50.63	58.45	63.44	73.19	84.03	78.00	82.20
4	29.37	36.67	50.08	59.12	70.23	54.40	68.20

Note. Mean data from children of ages 4–9 years (Gathercole, personal communication).

dygook (albeit phonotactically plausible gobbledygook), real subjects will decline to make any response. Wherever this cut-off point occurs, the model predicts that structural deletions (and insertions, see below) will increase in frequency as list length increases, especially for “mixed” lists (where adjacent items share few slots in the syllabic template) and where subjects are encouraged to guess rather than make “don’t know” responses.

Insertions

In the model, insertions occur when a template node ‘fires’ during the recall of the wrong syllable. This is often due to the slot being primed by input from nodes representing forthcoming syllables in the sequence. Thus they are more frequent in lists for which adjacent syllables use different slots as in:

/ʃim/, /wɒd/ → /ʃwim/, ...

A template node may also fire during the wrong syllable as a result of noise in its input. As noted under in the discussion of deletions, this is likely to occur as the strength of the nonspecific input, ϵ, increases for longer lists.

Insertions and deletions result largely from interactions between target syllables. These interactions produce more errors as a weight decay takes its toll on the syllables’ structural representations. This will occur not only as list length increases but also if decay is more rapid, as we hypothesize it is in young children (and patients with certain neurological disorders). The model is thus in agreement with data from Treiman (in press), which shows that, as memory span increases with age, there is a corresponding reduction in the proportions of errors involving the addition or deletion of phonemes.

Testable Predictions

The previous section summarized the effects of the syllable structure of list items on recall, and it is largely in this area that the model makes predictions that go beyond those which might be yielded by models which are unconstrained by linguistic considerations (e.g., Glasspool, 1994). Broadly speaking, the current model predicts that repetition of a particular stimulus, or serial recall of a particular list, will be affected by the structure of the syllables involved, as well as the number of phonemes and syllables. In order to examine these effects, stimuli must be constructed with regard to their syllabic structure (which slots they use). Lists in which the syllable structure alternates between items (e.g., lists of the following form, where the subscript denotes the slot in the template associated with each phoneme, C_1V, VC_5, C_1V, VC_5, C_1V, VC_5) can be compared with lists of the same length, in which syllable structures are blocked within lists (C_1V, C_1V, C_1V, VC_5, VC_5). In alternating lists, there is less overlap between the structure of adjacent items. We would therefore expect:

1. Insertions and deletions to be relatively more frequent than for blocked lists (where adjacent items share the same structure except at the edge of a block).
2. Contextual substitutions to be less frequent.

Because the two effects tend to cancel each other out, there may be little difference in the overall error rates which apply for the two conditions.

DISCUSSION

Extending the Model: Short-Term Memory and the Lexicon

The performance of the current model shows many of the characteristics of human subjects engaged in short-term memory tasks involving nonwords. It does this without any contribution of long-term lexical–phonological knowledge. However, studies showing effects of wordlikeness on nonword repetition (Gathercole, Willis, Emslie, & Baddeley, 1991) and familiarity effects in serial recall (Hulme et al., 1991) demonstrate that there is a contribution of long-term knowledge to short-term phonological memory. In order to account for some of the phonological regularities in errors of short-term memory for non-words, the current model postulates phonological retrieval processes common to speech production and short term memory. This development makes it possible to extend the model to address, within a unified computational model, the way in which short- and long-term phonological memory systems interact.

The current architecture may be extended to include long-term representations of lexical–phonological information, as outlined in Fig. 9 (cf. Fig 1). The phonological information required to produce familiar words is represented in the same way as that used in the simulations reported here, except that the connections which hold the information are *permanent* rather than temporary. When learning an unfamiliar word, nodes in the familiar-syllable group are partially activated by wordlike items in the input stream. Each becomes associated with a particular context state, and thus is reactivated at the appropriate serial position at recall, in much the same way as word nodes are in the articulatory loop model of Burgess and Hitch (1992). The permanent connections provide input to the template and phoneme groups, where it augments the input from the temporary pathways.

Note that while words and nonwords are represented separately, there is considerable overlap between the pathways. The same syllabic template is employed in the representation of both words and nonwords, and both word and nonword pathways activate the same phonological output nodes. The permanent pathways can be used in isolation to activate the phoneme group in speech production. In this mode of operation, the suggested model will function in a similar manner to Dell's spreading activation model of sentence production (Dell, 1988) and would be expected to produce a similar pattern of errors.

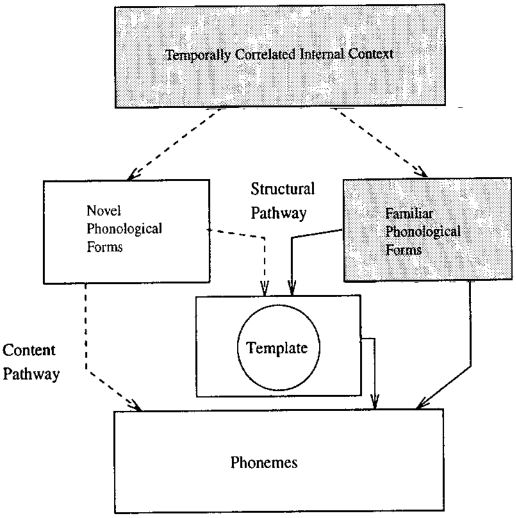

FIG. 9. Extending the existing architecture to include the representation of long-term lexical–phonological knowledge. The shaded area represent structures not included in the current implementation. Note that dashed lines represent pathways made up of temporary weights, the solid lines represent connections which do not decay over time.

A global increase in the rate of decay in the temporary weights of the network will cause more phonological misorderings for unfamiliar materials, as is the case in the simulation described above. Familiar materials (words) will not be similarly affected, because the connections in this pathway are permanent. However, a general reduction in memory span is to be expected because the connections between the internal context units and syllable units representing both words and nonwords are temporary ones (as in Burgess & Hitch, 1992). The suggestion that a global manipulation differentially affects various aspects of the model's performance provides a more parsimonious explanation of the nature and effects of damage than an approach which assumes that observed dissociations reflect the lesioning of independent functional modules (e.g., Caramazza et al., 1986).

It is further suggested that the permanent weights are established by means of an iterative procedure which exploits the temporary pathways. This could be achieved, for example, by rehearsing the material held in the temporary weights in short-term store, while activating a syllable unit in the long-term phonological store (an appropriate rehearsal mechanism is used by Burgess & Hitch, 1992 to refresh the temporary weights in their model). Thus the acquisition of new lexical–phonological knowledge will depend upon the effective operation of the short-term component of the model.

The Model in Relation to Other Work

The performance of the present model cannot be directly compared to any alternative account for the simple reason that there is no other model which covers precisely the same empirical ground. As discussed under Theoretical Background, we have built the model mainly on previous neural network models of serial order and the articulatory loop (Burgess & Hitch, 1992; Houghton, 1990) and of speech production (Dell, 1986, 1988). The experimental work discussed earlier does not make reference to these works, indeed, discussion of the data considered here generally makes no reference to any computationally precise model of serial recall or speech production. It is possible, however, to compare the model's handling of single-trial sequence learning, and of phonological retrieval, with models which share these capabilities.

Models of Serial Order and Short-Term Memory

One of the most important theoretical constraints on a short-term memory model is that it must be capable of single-trial sequence learning. Many models of serial order can be ruled out because they do not satisfy this constraint. Models employing variants of the backpropagation learning algorithm cannot offer a satisfactory account of short-term memory phenomena because they require multiple exposures to a sequence to learn it. For this reason, models of this kind, though capable of serial output, cannot be used to explain serial behavior in short-term memory tasks.

As far as models of short-term memory are concerned, surprisingly few are capable of single-trial serial learning and even fewer address the problem of how new phonological forms may be learned. Houghton (1990) put forward a model which is capable of single-trial learning, but in practice can only learn very short sequences of phonemes in a single exposure. Glasspool (1994) extended the Burgess and Hitch (1992) model to code novel forms of arbitrary length in a single trial. These models are not linguistically constrained. They have no mechanism for dividing a sequence into syllables, or of representing linguistic structure; phonological sequences are essentially treated as streams of undifferentiated symbols. Thus, they predict no organization at the level of the syllable. Ordering errors, if they occur, will very likely cross syllable boundaries and will not conform to any phonological principle: a sequence like /ger/, vaŋ/, kus/ might easily be recalled as /grev/, /ŋaku/, /s/. It is extremely doubtful whether such unpronouncable errors ever occur. We therefore do not consider these accounts as competing with our own, and they have never been applied to the same data. The same criticisms could be applied to any linguistically unconstrained model; without some form of representation of syllable structure, it is very unlikely that an implemented model would produce errors of the character described by Treiman and Danis (1988). The current model accounts for the phonological regularities in recall errors and is in other respects consistent with the B&H model, which provides a good basis for the understanding of many of the important phenomena associated with serial recall, but does not explain how novel forms can be stored and retrieved.

When Ellis (1980) first suggested that phonological errors in speech and short-term memory originated at the same locus, the lack of a computational model of verbal short-term memory and phonological retrieval made it difficult to see exactly where such an overlap might occur. The current model clarifies the situation somewhat: if syllable structure is implicated in the coding and retrieval of phonological information in short-term memory, it only really makes sense if the same processes are involved in spontaneous speech production; common retrieval processes are responsible for ordering segments for output, whether they are based on short- or long-term representations. Therefore, proposed structures must be consistent with speech error data and with the prevailing psycholinguistic accounts of phonological retrieval which account for them.

Models of Speech Production

In comparing the current model to models of speech production, its most notable difference is its capacity for single-trial learning. Many models in the area do not address the issue of how phonological forms are learned at all. Those that do (e.g., Jordan, 1986; Dell, Juliano, & Govindjee, 1993) have generally used backpropagation learning or related algorithms which require multiple exposures to a sequence in order to learn it. In the current model, the single trial learning of a target syllable's serial structure is made possible by associating its constituent phonemes with a simple reproducible sequence representing well-learned constraints on syllable structure. In contrast to Dell et al. (1993), our approach thus requires that structural (ordering) information is separated from content information. It is not at all clear that a model which did not separate structural and content information could account for phonological constraints in short-term memory errors.

As a model of phonological retrieval however, the current model has a good deal in common with many speech production models (e.g., Shattuck-Hufnagel, 1979; Stemberger, 1985; MacKay, 1987; Dell, 1986, 1988). The sequence of sounds to be output is represented at more than one level, and selection at each level is on the basis of competition between concurrently active elements. Most linguistic accounts also share the assumption put forward here that retrieval involves two processes: the specification of a structural frame for the target utterance, and selection of the phonological contents of that frame.

Syllable Structure

The current model employs a single generalized syllable schema, which is dynamically reconfigured to represent specific syllable structures. This is clearly more parsimonious than a model that proposed that there is a separate frame for each legal syllable structure in the language. It also provides a more appealing account of additions (and deletions). The structural representations of adjacent syllables can interact and affect which slots are accessed (e.g., leading to errors like /ʃim/, /wɒd → /ʃwim/, ... or, say, "lead pipe → plead pipe"). It might otherwise appear that "the wrong frame" has been selected arbitrarily.

The particular form of the syllabic-template implemented here imposes a fairly minimal sonority-based scheme for the intra-syllabic ordering of output segments. It has a number of properties which make it useful for the modeling of verbal short-term memory. In particular, its cyclical character means that the structure of a series of syllables can be determined (and thus learned) *while it is being experienced.* It also allows retrieval of syllabic structure to be achieved without the postulation of top-down processes of much complexity. However, it is clearly something of a simplification and is occasionally at odds with linguistic theory; we believe that the benefits of postulating this simplified parsing structure outweigh the costs.

For example, the template parses syllables such that the number of phonemes in the rhyme is maximized. Linguistic accounts of syllabification, on the other hand, typically involve a maximal onset rule (e.g., Clements & Keyser, 1981), by which intervocalic consonants are assigned to the onset of the later syllable, wherever they are legal in that position.[6] Speech production models can be consistent with syllabification theory by simply specifying frames which have the desired properties. They are free to do so because, in the main, they do not attempt to explain how a particular structural representation becomes associated with a particular phonological form in the first place. in contrast, a short-term memory model must specify how the structural frame for a particular syllable is computed *as it is perceived.* It is difficult to see how this theoretical problem can be reconciled with a maximal onset rule, which would require access to a span of input: in order to implement a maximal onset rule, before one phoneme can be parsed, the identify of the next is often required. For example, the /l/ in "melon" is assigned to the second syllable, but the /l/ in "melting" must belong to the first, because the sequence /lt/ cannot occur in the onset of a syllable.

We feel that ultimately additional linguistic constraints on syllable structure (such as the maximal onset rule referred to above) can probably be reconciled with the kind of bottom-up "on-line" parser we have specified here. How-

---

[6] Empirically, results from short-term memory experiments show that this constraint is by no means absolute, and that a more flexible account of syllabification is required (Treiman, Straub, & Lavery, 1994).

ever this is likely to require a much more complex “syllable template,” which would considerably complicate the model. The current implementation allows the model to deal with continuous input, and its treatment of syllable structure does not result in any serious discrepancies with the short-term memory data.

The template proposed in the current model is an approximation. Whether or not one regards it as realistic, the central idea that a syllable frame or schema is used in the representation of novel stimuli in short-term memory explains what would otherwise be puzzling phenomena. The elaboration of the real underlying structure or structures requires more qualitative analyses of short-term memory errors, which offer the opportunity to study linguistic constraints on phonological retrieval under experimentally controlled conditions. We hope that the model put forward here will serve to stimulate such research.

Conclusion

The importance of short-term memory for novel words in the development of language is becoming increasingly apparent (see, e.g., Gathercole & Baddeley, 1989). In order to understand its developmental role, and its interactions with long-term memory, a computational model of this system will be useful. In the learning and recall of nonwords, linguistic constraints on short-term memory become much more apparent than they are when the stimuli are familiar words. This makes the area fertile ground for theoretical investigation—the opportunity exists for theories of speech production and short-term memory to be integrated in explaining phenomena associated with the task. In this model, we have taken the first steps toward such an integration.

APPENDIX 1: FORMAL SPECIFICATION OF THE MODEL

This appendix specifies the mathematical details of the model and gives the parameter values used in the simulations.

Activation

The activation level a_i of a node u_i is a sigmoidal function of its input I_i. Specifically,

$$a_i = \frac{1}{1 + e^{-\tau(I_i - 1)}} + \nu. \qquad (3)$$

The term v is a random variable (representing noise) drawn from a Gaussian distribution of an 0.0 and standard deviation σ (set to 0.065 in the reported simulations). The parameter τ determines the slope of the sigmoid, larger values giving steeper slopes. A value of $\tau = 5.0$ (giving a gentle slope) is used in all simulations for nodes which can be active in parallel (syllable and phoneme nodes). By contrast, it is desirable that nodes in the syllable template are either fully active or inactive. This “all-or-none” behavior is implemented by setting τ to a very high value (e.g., 100.0) for template nodes.

The net input, I_i, to a node u_i is the sum of the weighted activations of nodes connected to u_i, given by

$$I_i = \sum_i \max(0, a_j) w_{ji}, \qquad (4)$$

where a_j is the activation of a node connected to u_i, and w_{ji} is the weight on the link from u_j to u_i. Negative activation values are not propagated.

Learning

The strengths of the temporary connections are developed by exposure to a stimulus sequence according to Eq. (1), given in the main text. The strengths of the permanent connections are fixed throughout learning and recall at 1.0.

The learning rule is Hebbian in character, being unsupervised and dependent on the concurrent activation of connected nodes. During learning, nodes in the network are activated in response to the serial input as described in the main body of the text. Nodes are activated to a maximum value (1.0) for one timestep, with one important exception: a number of authors have noted that vowels are less prone than consonants to serial order errors (Ellis, 1980; Treiman & Danis, 1988; Patterson et al., 1994), and it has been suggested that this is because they are longer in duration. Vowels also have a greater acoustic intensity than consonants. These differences are likely to affect on the strength of the trace laid down by different speech sounds. To model this approximately, nodes representing vowels are maximally activated when they appear in the input stream, whereas the activation of nodes representing consonants is set to 0.75.

Recall

During recall, the units in the syllable layer are assumed to be activated by input from a CQ mechanism such as that described by Burgess and Hitch (1992). A competitive filter ensures that syllables that have already been produced and suppressed. Sequencing at the word/nonsense syllable level is assumed to operate correctly, so that during recall, the syllable unit (onset/rhyme) pair representing the current target output receives the most input (I_{max}). Upcoming syllables have less activation the further they occurred from the current target during learning. The activation of each syllable s_0 to s_n is given by equation 5, where s_i designates the syllable node associated with the i^{th} target syllable, and the current target is the j^{th}:

$$a_{si} = \begin{cases} f(I_{max}(1 - \frac{i-j}{r})) & \text{in the range } j \leqslant i < j + r \\ 0 & \text{outside it.} \end{cases} \tag{5}$$

In the simulations reported here I_{max} is set at 1.0, and r is 3.

In order to produce a response during recall, the syllabic template must be accessed such that the template nodes associated with the target syllable fire in series. It is proposed that this is achieved by applying an additional input of magnitude ϵ cyclically to each node in the template group in turn. Thus, the net input to a template node (designated u_k) becomes

$$I_k = \sum_i a_{s_i} w_{ik} + \epsilon. \tag{6}$$

The value of ϵ can be thought of (loosely) as the "energy" or "effort" required to drive the syllable template, irrespective of its content. This additional input should be of sufficient magnitude to cause a template node to fire if it is receiving input from the syllable unit representing the current target, but not otherwise. The optimal value of ϵ depends on the duration of the list, and on the values of other parameters in the model.

A template node, u_k fires if I_k is greater than 1.0. Clearly, if the value of ϵ is greater than 1.0, then all the template nodes will fire regardless of the input they receive from the syllable group. This is undesirable because most syllables will not involve all the positions in the syllable template. Ideally, I_k should only exceed 1.0 for template nodes representing slots used in the target syllable. However, the priming of syllable units representing syllables further on in the intended sequence means that nodes not used in the current target will fire if the value of ϵ is too great. In order to prevent this occurring, the value of ϵ must be such that with input from the primed syllable nodes (s_{j+1} thur s_{j+r-1}) I_k is less than 1.0. This gives an upper limit to the value of ϵ:

$$\epsilon_{max} = 1.0 - \sum_{i=j+1}^{j+r-1} a_{s_i} w_{s_i k}. \tag{7}$$

At the same time, it is important that the template node fires if it *is* connected to the target syllable unit. This places a lower limit on the value of ϵ, which must be sufficient to fire a template node u_k even if its input comes solely from the target syllable s_j:

$$\epsilon_{min} = 1.0 - a_{s_j} w_{s_j t}, \tag{8}$$

If it is assumed that both insertions and deletions are equally unfavourable, the optimal value of ϵ is half way between these limits. For a given set of parameters, the individual terms in Eqs. (7) and (8) can be calculated for a notional template node connected to the appropriate notional syllable units, and an approximation of the optimal value of ϵ determined: weights are assumed to have an initial strength of α. When recalling the j^{th} syllable, the decay factor $\delta_i - j$ on connections from s_i, activated in response to the i^{th} input syllable is given by:

$$\delta_{i-j} = 0.5^{\frac{l-(i-j)p}{h}}, \tag{9}$$

where l is the list duration, and p is the duration of pauses between syllables and $i \geqslant j$. For $r =$

3 and $\alpha = 1$, the optimal value of ϵ is approximated by:

$$\epsilon_{opt} \approx \frac{2 - f(I_{max})\delta_0 - f\left(\frac{2I_{max}}{3}\right)\delta_1 - f\left(\frac{I_{max}}{3}\right)\delta_2}{2} \qquad (10)$$

Since the weights from syllable units to template nodes decay over time, the optimal value of ϵ is dependent on the context in which recall occurs (e.g., $\epsilon_{opt} \approx 0.84$ in the simulations of Treiman and Danis' stimuli; $\epsilon_{opt} \approx 0.70$ for the three-syllable nonword repetition stimuli). If a relatively short period has elapsed between learning and recall, weights from the syllable group to the context group will have decayed little, and will provide plenty of input to the template nodes. Relatively little extra input will be needed to make the primed nodes fire. When the interval between learning and recall is longer, more external activation will be necessary to augment the decaying input from the syllable nodes. As time passes, the strength of the syllable-template connections tend toward zero and both upper and lower limits on ϵ converge on 1.0. As ϵ increases the slots that fire at recall will eventually be determined solely by the noise in the activation of the syllable units. Ultimately any attempt at recall is almost certainly doomed to failure. Rather than attempt to model "don't know" responses, we assume that some attempt is made at recall for each syllable presented. As the temporary weights in the network decay, these responses will take on the character of guesses. However, initially, they will at least maintain the structural properties, and often some of the phonemic content of the target syllable.

On the basis of the activations of template and syllable units, and the strengths of the weights in the network, the activation of the each of the phoneme nodes is calculated in the normal way. Nodes compete to be output on the basis of their activations. The phoneme represented by the most active mode will be output providing its activation exceeds a threshold (to prevent noise resulting in continuous output). The threshold was set at 0.5 in the simulations.

Rate of Presentation and Recall

It is important in any model involving trace decay to be clear about the relationship between rates of presentation and recall. If the rates differ, part of the list will experience more trace decay than others, assuming (as we do here) that the rate of decay is constant throughout learning and recall. The current implementation of the model runs in simulated real time, that is in steps which are assigned a meaningful duration. This approach has the advantage that the temporal parameters of the model, syllable duration, pause duration and temporary weight half-life can be expressed in real units. In the current implementation, the duration of an utterance is determined by the number of syllables it is made up of, and, as a simplification, we assume that all syllables have the same duration. During presentation, this constant duration is divided into timesteps of equal duration, one for each phoneme to be presented, During recall, the syllable's duration is divided into five equal timesteps during which each of the slots in the syllabic template is accessed.

In the reported simulations, syllable duration was set at 0.4 s. In modeling nonword serial recall, a presentation rate of 1.0 syllables per second was used, as in Treiman and Danis (1988). Thus a pause of 0.6 s was allowed between presented syllables during which weight decay continued. In simulations of nonword repetition, there were no pauses during learning or recall.

APPENDIX 2: NETWORK FOR INPUT PROCESSING

The model we have described relies heavily on a capacity to track the incoming stream of phonological information and parse it using a cyclical syllabic template. This must be achieved using information available in the input stream and contextual information from the "wheel" itself. In the following section we describe a sequential connectionist machine which performs the parsing process we have described. In this network, the phonological constraints are acquired through exposure to words. A supervised learning algorithm is used to adjust these permanent weights to produce an expected series of output patterns in the template nodes for each word presented. These well-learned phonological constraints generalize to unfamiliar forms.

The following architecture was adopted (see Fig. 10); each template node receives input from all of the phoneme nodes, and from each of the nodes in the template group, including itself. Lateral connections at the level of the template group are necessary, because the wheel must operate *cyclically* on the input, and thus requires information about its prior states. Because the activation state of the template group at one timestep influences its subsequent states, the structure is recurrent. Networks like these can be trained using a version of the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986, pp. 355–360). A series of patterns is presented to the network, producing a sequence of patterns of activation in the output units (in this case the template nodes). As in the standard backpropagation of error algorithm, each output pattern can be compared with a training pattern, which represents the desired or expected outcome of processing the current input pattern in its particular temporal context. For each node, the difference between the observed and expected outcomes is computed, for the current state of the network and for its state during previous timesteps. The net contribution of each *current* weight to the total error (summed over time) is calculated. Each weight can then be adjusted such that this contribution to the error is reduced by a small amount. Gradually, the weights converge on values which minimize the error for each pattern in training set.

To train the network, a set of 803 phonetically encoded words was constructed using the Oxford Psycholinguistic Database (Quinlan, 1992). Each word had an age-of-acquisition less than 8 years (Gilhooly & Logie, 1980). The list contained a fair proportion of polysyllabic words, as well as a large number of simple monosyllables. The same representation of phonology was used as in the short-term memory model. Each input pattern was a vector with one non-zero element standing for the currently presented phoneme. Training vectors were prepared using the same procedure as in the short-term memory model, that is for each phoneme in the input stream, the template node standing for the next available clockwise slot was expected to be maximally active.

The complete sequence of 4133 input vectors was presented to the network 50 times, while learning took place. After this period, the network tracked the input stream very much as described under the heading "The syllabic template." As each phoneme appeared, the next available clockwise matching slot in the template group became very active, while the other nodes remained inactive. Table 9 shows the activations of the five nodes during the presentation of the (unlearned) sequence/slædiŋ/. As can be seen, the mechanism parses the novel input into the syllables /slæd/ and /iŋ/.

The connections which produce this behavior are represented schematically in Fig. 11. White boxes represent inhibitory weights, black boxes

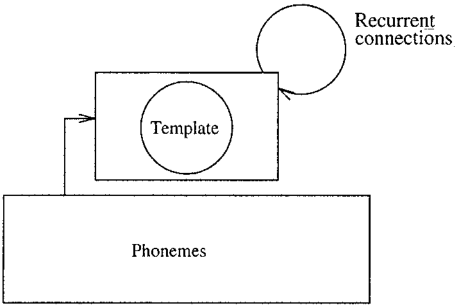

FIG. 10. Architecture for input processes. All connections are permanent.

TABLE 9
TABLE SHOWING ACTIVATIONS OF TEMPLATE NODES DURING PRESENTATION OF AN UNFAMILIAR PHONOLOGICAL SEQUENCE

	Template "slot"				
Phoneme	1	2	3	4	5
s	0.950	0.009	0.004	0.000	0.029
l	0.007	0.967	0.030	0.006	0.000
æ	0.005	0.003	0.998	0.001	0.000
d	0.013	0.000	0.007	0.021	0.983
ɪ	0.026	0.027	0.965	0.000	0.000
ŋ	0.000	0.000	0.044	0.949	0.033

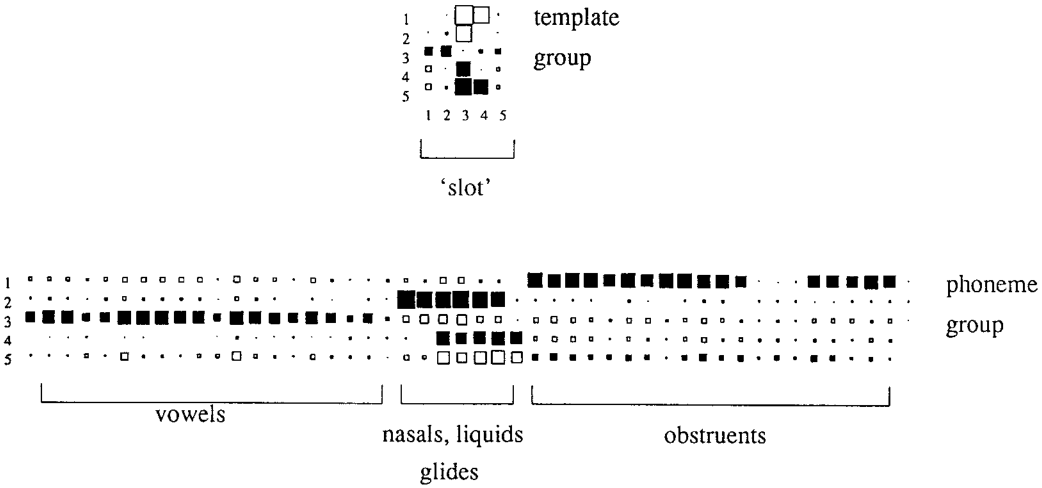

FIG. 11. Connections to template nodes after training.

excitatory ones. The strength of the connection is indicated by the size of the box. The lower part of the figure shows the bottom-up connections from phoneme to template group; the upper part shows the recurrent connections between template nodes. In each case the top row of the matrix represents connections to the first template position, the second row connections to the second slot, and so on.

The pattern of connections from the phoneme layer is much as we might expect. If in the training input, the phoneme is associated with a particular slot, an excitatory connection to the appropriate template node is developed. There is a tendence for phonemes to inhibit slots with which they are not associated. This is particularly true of the nasals, liquids and glides. It is also notable that the obstruent segments tend to activate the pre-vocalic slot more than the post-vocalic one.

The recurrent connections in the template group are interesting. The positions associated with pre-vocalic consonants (slots 1 and 2) tend to activate the vowel node, while inhibiting the post-vocalic consonantal position (4 and 5). The vowel position (slot 3) strongly inhibits the pre-vocalic consonantal (onset) positions, and strongly excites the post-vocalic consonantal positions. This explains the asymmetry in the phoneme to template weights for consonants. Extra bottom-up input from the phoneme group is required to overcome the inhibition of the pre-vocalic slot. Weights from the template node representing slot 4 (which is associated with post-vocalic nasals, liquids, and glides) seem to be principally concerned with the interpretation of a following obstruent segment, tending to inhibit the template node associated with slot 1 and excite the one associated with slot 5. The final template position tends to activate slot 3—the vowel position, while inhibiting the postvocalic consonantal positions.

We have a number of reservations about the learning algorithm, and cannot claim it is a realistic model of how such phonological constraints are acquired. In particular, it is not clear where the teaching input that the algorithm requires could originate. We tentatively suggest that it may be possible to derive a measure of within-syllable position of "syllabic phase" directly from changes in the acoustic energy of the signal over time. This kind of information can certainly be used to locate syllable boundaries (Mermelstein, 1975; Hunt, 1993) and the "beats" or perceptual centers of isolated syllables (Marcus, 1981; Scott, 1993). Our own investigations in the area are, as yet, at a very preliminary stage. For the moment, we put forward the recurrent network as a demonstration that in principle, a connectionist machine capable of performing the cyclical parsing we have described can be specified, and that constraints learned through exposure to words generalize to novel stimuli.

REFERENCES

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

BISIACCHI, P. S., CIPOLOTTIS, L., & DENES, G. (1989). Impairments in processing meaningless verbal material in several modalities: The relationship between short-term memory and phonological skills. *Quarterly Journal of Experimental Psychology,* **41A,** 292–320.

BRADY, S., SHANKWEILER, D., & MANN, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology,* **35,** 345–367.

BURGESS, N. (1995). A solvable connectionist model of immediate recall of ordered lists. In G. Tesauro, D. Touretzky, & T. K. Leen (Eds.), *Neural information processing systems.* Cambridge: MIT Press.

BURGESS, N., & HITCH, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language,* **31**(4), 429–460.

CAMPBELL, R., & BUTTERWORTH, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *Quarterly Journal of Experimental Psychology,* **37A,** 435–475.

CARAMAZZA, A., MICELI, G., & VILLA, G. (1986). The role of (output) phonology in reading writing and repetition. *Cognitive Neuropsychology,* **3,** 37–76.

CHI, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits. *Memory and Cognition,* **4,** 599–572.

CLEMENTS, G. N., & KEYSER, S. J. (1981). *A three-tiered theory of the syllable.* Technical report, Massachusetts Institute of Technology.

DELL, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review,* **93**(3), 283–321.

DELL, G. S. (1988). The retrieval of phonological forms in production: Test of predictions from a connectionist model. *Journal of Memory and Language,* **27,** 124–142.

DELL, G., JULIANO, C., & GOVINDJEE, A. (1993). Structure and content in language production: A theory of frame constraints in phonological errors. *Cognitive Science,* **17,** 149–195.

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 organisation in memory. In *Coding processes in human memory* (pp. 161–190). Washington, DC: Winson.

FUDGE, E. C. (1969). Syllables. *Journal of Linguistics,* **5,** 193–320.

GATHERCOLE, S. E., & ADAMS, A. (1993). Phonological working memory in very young children. *Developmental Psychology,* **29**(4), 770–778.

GATHERCOLE, S. E., & BADDELEY, A. D. (1989). Evaluation of phonological S T M in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language,* **28,** 200–213.

GATHERCOLE, S. E., WILLIS, C. S., BADDELEY, A. D., & EMSLIE, H. in press. The children's test of nonword repetition: A developmental indicator of language ability.

GATHERCOLE, S. E., WILLIS, C. S., EMSLIE, H., & BADDELEY, A. (1991). The influences of number of syllables and word-likeness on children's repetition of nonwords. *Applied Psycholinguistics,* **12,** 349–367.

GILHOOLY, K. J., & LOGIE, R. H. (1980). Age of acquisition, imagery, familiarity and ambiguity measures for 1944 words. *Behavioral Research Methods and Instrumentation.* **12,** 395–427.

GLASSPOOL, D. W. (1994). Competitive queuing and the articulatory loop. In *Connectionist models of memory and language.* UCL Press.

GOSWAMI, U., & BRYANT, P. (1990). *Phonological skills and learning to read.* London: Erlbaum.

HOUGHTON, G. (1990). The problem of serial order: A neural network memory of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation.* London: Academic Press.

HOUGHTON, G., GLASSPOOL, D. W., & SHALLICE, T. (1994). Spelling and serial recall: insights from a competitive queueing memory. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of normal and disturbed spelling.* Chichester: Wiley.

HOUGHTON, G., & HARTLEY, T. (in press). Parallel models of serial behaviour: Lashley revisited. *Psyche.*

HULME, C., MAUGHAN, S., & BROWN, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language,* **30,** 685–701.

HULME, C., & SNOWLING, M. (1992). Deficits in output phonology: An explanation of reading failure? *Cognitive Neuropsychology,* **9**(1), 47–72.

HUNT, A. (1993). Recurrent neural networks for syllabification. *Speech Communication,* **13,** 323–332.

JORDAN, M. I. (1986). *Serial order: A parallel distributed processing approach.* Technical Report, 8604, Institute for Cognitive Science, University of California, San Diego.

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

MACKAY, D. G. (1987). *The organisation of perception and action.* New York: Springer-Verlag.

MARCUS, S. M. (1981). Acoustic determinants of perceptual centre (P-center) location. *Perception and Psychophysics,* **30,** 247–256.

MARTIN, N., SAFFRAN, E. M., DELL, G. S., & SCHWARTZ,

M. F. (1994). Origins of paraphasias in deep dyslexia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval.

MERMELSTEIN, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America,* **58**(4), 880–883.

MEYER, A. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language,* **30,** 69–89.

PATTERSON, K., GRAHAM, N., & HODGES, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience,* **6**(1), 57–69.

QUINLAN, P. T. (1992). *The oxford psycholinguistic database.*

RUMELHART, D. E., HINTON, G., & WILLIAMS, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing, volume 1: Foundations.* Boston, MA: MIT press.

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

SCOTT, S. K. (1993). *Perceptual centres in speech—An acoustic analysis.* Ph.D. thesis, University College London.

SELKIRK, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Orehrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teachers and students.* Cambridge: MIT Press.

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

STEMBERGER, J. P. (1985). An interactive activation model of language production. In Ellis, A. (Ed.), *Progress in the psychology of language* (pp. 143–186). Hillsdale, NJ: Erlbaum.

TREIMAN, R. in press. Errors in short-term memory for speech: A developmental study.

TREIMAN, R. (1986). The division between onsets and rhymes in english syllables. *Journal of Memory and Language,* **25,** 476–491.

TREIMAN, R., & DANIS, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **14**(1), 145–152.

TREIMAN, R., STRAUB, K., & LAVERY, P. (1994). Syllabification of bisyllabic nonwords: evidence from short-term memory errors. *Language and Speech,* **37**(1), 45–60.

(Received May 24, 1994)

(Revision received March 3, 1995)