

# The Role of Syllable Structure in Verbal Short-Term Memory

Tom Hartley

University College London

September 20, 1995

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

---

## Abstract

Remembering the sound of a new word when it is first encountered is an important skill which plays a critical role in the development of vocabulary (Gathercole & Baddeley, 1989), yet the mechanisms underlying this form of verbal short-term memory are not well understood. Errors in the repetition and serial recall of nonwords indicate that structural properties of the syllable are represented in short-term memory, but existing accounts of serial learning and recall do not incorporate any representation of linguistic structure. Models of speech production implicate syllable structure in the representation of phonological form, but do not explain how such representations are acquired. This thesis draws together theories of speech production and serial memory to develop a computational model of nonword repetition based on the novel idea that short-term memory for the serial order of a sequence of speech sounds is constrained by a syllabic template. The results of simulations using the model are presented and compared with experimental findings concerning short-term memory for nonwords. The interaction of short- and long-term phonological memory systems and the acquisition of vocabulary are discussed in terms of the model. The model is evaluated in comparison with other contemporary theories.

---

# Acknowledgements

The research described in this thesis was supported by a grant from the SERC.

Rebecca Treiman and Sue Gathercole generously provided supplementary information and data from some of their experimental studies which helped enormously in developing the model described in chapter 5

I would like to thank my supervisor, George Houghton, for helping me find a subject worth investigating, for giving me the freedom to explore it and for providing encouragement and inspiration. Many others at UCL helped me in one way or another, by providing inspiration, advice and support. I would particularly like to thank Stevie Sackin, David Glasspool, Paul Hibbard, John Draper, Neil Burgess, Willie Curran, Colin Clifford, Keith Langley, Rick Cooper, and Jonny Farrington.

There are many people whose humour and friendship helped me to complete this project: Tom Manly, my brother Olly, James Taylor, Jenny Porter and The Crayons all helped me carry on. Special thanks are due to Sophie Scott, who is a great friend. She encouraged me to come to UCL in the first place, set a good example, helped me get my act together and made working here a lot of fun.

Without Rachel Harris, I could not have even begun this thesis. I cannot thank her enough for the constant love and support she has given me. I would also like to thank all my family especially my mum, Di Beattie, my dad, Ian Hartley, and my stepmother Jennifer Hartley for all the encouragement they have given me.

This thesis is for my grandparents. Thankyou.

# Contents

<b>1</b>	<b>Introduction</b>	<b>18</b>
1	General introduction . . . . .	18
2	Short-term memory . . . . .	19
2.1	Functions of short-term memory . . . . .	21
2.2	The role of speech in short-term memory . . . . .	24
2.3	Serial order and short-term memory . . . . .	25
2.4	Errors in speech and short-term memory . . . . .	26
3	Overview . . . . .	28
<b>2</b>	<b>The problem of serial order: general issues and models</b>	<b>29</b>
1	'Buffers' . . . . .	31
2	Associative chaining . . . . .	33
2.1	Representing repeated items and multiple sequences . . . . .	34

# CONTENTS

---

2.2	Coarticulation and Anticipation . . . . .	37
2.3	More recent 'chaining' models . . . . .	39
3	Parallel models of serial order . . . . .	40
3.1	I-I models . . . . .	43
4	P-I models . . . . .	51
4.1	Competitive queuing . . . . .	54
5	Summary and Conclusions . . . . .	63
<b>3</b>	<b>Empirical background: serial order and memory for speech</b>	<b>64</b>
1	Stimulus familiarity and recall . . . . .	64
2	Phonological constraints on recall . . . . .	72
3	Constraints on speech errors . . . . .	74
3.1	Errors in spontaneous speech . . . . .	75
3.2	Experimentally elicited slips . . . . .	79
4	Summary . . . . .	80
<b>4</b>	<b>Modelling constraints on phonological retrieval: a cyclical syllable template</b>	<b>82</b>
1	Phonological slips, STM, and CQ models . . . . .	83

## CONTENTS

---

2	Psycholinguistic models of phonological retrieval . . . . .	90
3	General constraints on syllabic phonology . . . . .	95
4	A syllabic template . . . . .	98
4.1	Parsing the syllable . . . . .	98
4.2	Parsing continuous 'speech' . . . . .	100
<b>5</b>	<b>A new model of short-term memory</b>	<b>103</b>
1	Description of the model . . . . .	103
1.1	Architecture. . . . .	104
1.2	Learning and Recall . . . . .	107
1.3	Recall . . . . .	111
2	Simulations . . . . .	115
2.1	Quantitative measures of performance . . . . .	116
2.2	Qualitative analyses . . . . .	118
2.3	Modelling a damaged/developing store . . . . .	126
2.4	Error mechanisms . . . . .	129
2.5	Syllable structure and errors . . . . .	136
2.6	Testable predictions . . . . .	139

---

# CONTENTS

---

<b>6</b>	<b>Future directions and conclusions</b>	<b>141</b>
1	Extending the model: short-term memory and the lexicon . . . . .	141
2	The model in relation to other work . . . . .	146
2.1	Models of serial order and short-term memory . . . . .	147
2.2	Models of speech production . . . . .	150
2.3	Syllable structure . . . . .	151
3	Conclusions . . . . .	153
<b>A</b>	<b>Coding used in phonological simulations</b>	<b>155</b>
1	Vowels . . . . .	156
2	Consonants . . . . .	157
<b>B</b>	<b>Technical appendix to chapter 5</b>	<b>158</b>
1	Activation . . . . .	158
2	Learning . . . . .	159
3	Recall . . . . .	160
4	Rate of presentation and recall . . . . .	163
<b>C</b>	<b>Exception words</b>	<b>165</b>

# CONTENTS

---

<b>D Dynamics of the syllabic template</b>	<b>169</b>
1 Dynamics of the syllabic template . . . . .	169
1.1 Capturing syllable structure in a recurrent network . . . . .	170
1.2 Determining syllabic phase . . . . .	175
2 Formal specification of the syllabic phase algorithm . . . . .	182
 <b>Bibliography</b>	 <b>185</b>



# List of Figures

- 2.1 *Representation of a sequence in the form of a chain of item-item associations. . . . . 34*
- 2.2 *a) Associative chaining has difficulty in accounting for the representation of sequences containing repeats. It is not clear which of the associative links from “E” should be followed. b) If a single associative structure is to represent multiple sequences, there is no mechanism to select those links which must be followed to generate a particular sequence. Here phonemic representations of cat, tack and act would interfere with one another (different types of arrows are used to show the links needed for each word). c) and d) Wickelgren’s context-specific coding overcomes these difficulties, but only at the expense of using entirely different tokens to represent instances of the same type . . . . 35*
- 2.3 *Structure of part of a typical neural network. A node  $j$  has a number of afferent connections. The weighted sum of these inputs ( $i_j = \sum_i w_{ij} o_i$ ) is used to determine  $j$ ’s activation ( $a_j = f(I_j)$ ) . The node’s output is a function ( $g$ ) of its activation. Each node’s efferent connections are weighted, and weights can be changed systematically so that the network learns to produce an adaptive response to its inputs. . . . . 42*

## LIST OF FIGURES

---

- 2.4 *Estes' (1972) scheme for the representation of serially ordered actions. For each sequence (e.g., "C,H,A,I,N"), there is a node which activates its nodes standing for its constituent items via excitatory links to each (solid arrows). Item nodes representing actions later in the sequence receive inhibitory input (dashed arrows) from nodes representing earlier items – for clarity, only the lateral connections from "H" are shown here; the later an item is in the list, the more items precede it. Items nearer the end of the sequence receive most inhibition, thus when the item nodes are activated in parallel by input from the sequence nodes, an activation gradient is set up. . . . . 45*
- 2.5 *Jordan (1986) and Elman (1990) have put forward similar schemes for the learning and representation of serial order in connectionist networks. In each case fixed recurrent connections (solid arrows) provide a time-varying context at the level of the input units. Modifiable feedforward connections (dashed arrows) are adjusted using supervised learning algorithms. . . . . 47*
- 2.6 *Grossberg's (1978) model uses weights of varying strength between sequence and item nodes to set up an activation gradient, which is used to control serial behaviour. . . . . 51*
- 2.7 *Sequences coded using Grossberg's scheme (1978) do not interfere in the way that sequences coded as associative chains do. Here, for example, the weights used to represent the sequences "A,C,T", "T, R, A, P" and "P, A, T" do not conflict with one another. . . . . 53*
-

- 2.8 *In competitive queuing (CQ) models, a time-varying control signal (or context) is used to encode serial position. Generally the signal, represented here by the shaded bar, changes gradually over time (from left to right), such that nearby states are positively correlated. Each item in a sequence is associated with a particular state of the control signal. Recovery of the control signal can be used to activate the items in sequence, but its temporally-autocorrelated nature results in competition between responses, especially if they are close together in the target sequence. . . . . 54*
- 2.9 *The representation of sequences containing repeats is not problematic for CQ models, since each instance of an item (e.g., each “E” in “E, V, E, R, Y”) can be associated with a distinct state of the control signal. 54*
- 2.10 *Houghton’s competitive queuing architecture (1990). Serial order is information is represented in a distributed fashion in the learned weights (a) between sequence and item nodes. A temporally-modulated activation gradient is established during recall by ‘replaying’ the control signal experienced during learning. Item nodes feed their activation forward to the competitive filter via one-to-one excitatory connections (b). Lateral inhibition (c) between filter nodes suppresses all but the most active item node, and allows that item to be output. The selected item is then suppressed by means of one-to-one feedback connections from filter to item nodes (d). . . . . 56*

## LIST OF FIGURES

---

- 2.11 *The changing activation of a set of item nodes in a Competitive Queuing model similar to that described by Houghton (1990). a) Recall of a short non-repeating sequence ('cliff'). b) Recall of a sequence containing a repeated item ('click') showing the recovery from suppression of the item node representing /k/. . . . . 57*
- 3.1 *The robustness of the short-term representation of the form of a familiar item is enhanced by the availability of a long-term representation. This long-term knowledge cannot be exploited in the representation of novel forms. . . . . 66*
- 4.1 *The sonority principle: a) each phonological segment is associated with a particular sonority value, reflecting its energy relative to other segments (based on a diagram from Ladefoged, 1993); b) in a well-formed syllable, segments are usually organised such that sonority rises to a single vocalic peak, and declines thereafter. . . . . 96*
- 4.2 *English syllable structure as described by Fudge (1969) . . . . . 97*
- 4.3 *The structure cyclical syllabic template used in the model described in chapter 5. . . . . 99*
- 5.1 *An outline of the model's structure . . . . . 105*
- 5.2 *A more detailed diagram of the model's structure showing nodes and connections. The dashed lines represent temporary weights, the solid lines permanent connections. . . . . 106*
-

## LIST OF FIGURES

---

- 5.3 *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. 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. . . . . 110*
- 5.4 *Diagram illustrating activation states of the network during recall of a nonword. As in figure 5.3 the degree of activation is represented by the density of the shading. . . . . 112*
- 5.5 *Graph showing proportions of correct responses in simulations of nonword repetition for items of varying length 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. . . 116*
- 5.6 *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,  $r^2 = 0.94$ ) 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. . 127*
-

## LIST OF FIGURES

---

- 6.1 *Extending the existing architecture to include the representation of long-term lexical-phonological knowledge. The shaded areas represent structures not implemented in the version of the model described in chapter 5. Note that dashed lines represent pathways made up of temporary weights, the solid lines represent connections which do not decay over time. . . . . 142*
- 6.2 *This figure (cf. figure 5.2) shows in detail how a familiar monosyllable ("barn") would be represented in the phonological lexicon. Note that all the connections are permanent. . . . . 144*
- 6.3 *An alternative architecture does not explicitly differentiate between nodes used to represent familiar and unfamiliar phonological forms. Instead, weights are subject to dual potentiation, having both permanent and temporary components (the adjacent dotted and solid lines arrows represent such dual weights). With repeated use, the permanent component gradually builds up. However it is still necessary to be able to distinguish unassigned nodes (with no permanent lexical association) so that they can be recruited to become associated with new words when they are first encountered. . . . . 145*
- D.1 *Architecture for a neural network capable of parsing syllables on-line according to the scheme developed in chapter 4 . . . . . 170*
- D.2 *Diagram showing weights developed by the recurrent network following repeated exposure to a set of English words. The size of the square represents the strength of the connection. Black squares represent excitatory connections, white squares inhibitory ones. . . . . 173*
- D.3 *The concept of phase as applied to a hypothetical syllabic cycle. . . . . 175*
-

- D.4 *A schematic representation of the temporal filtering of an ideal signal (a periodic one with frequency  $f$ ). The quadrature filter (represented on the upper axes) can be used to recover a phase description . . . . . 178*
- D.5 *For an ideal signal, the output of the filter ( $R$ ) traces a circular path in the complex plane. The phase of the signal ( $\phi$ ) is the angle  $R$  makes with the real axis at any time. . . . . 179*
- D.6 *Output from an implementation of the syllabic phase filter. The top graph shows the speech waveform. Beneath it is the phase output of the filter, giving a real-valued measure of within-syllable position for each frame of the sample. Beneath the phase output, the theoretical syllable boundaries are marked. The bottom graph shows the amplitude of the filters output. Beneath the graphs an approximate phonetic transcription is shown. . . . . 181*

# List of Tables

- 4.1 *The proportion of responses showing various consonant-vowel structures in simulations of serial recall of lists of six monosyllabic nonwords.* 89
- 4.2 *The proportion of errors produced by Glasspool's model with different phonemic lengths. Data from human subjects performing a similar task are shown for comparison (Treiman, in press) . . . . .* 89
- 4.3 *A continuous stream of phonemes is parsed into syllable-sized chunks by assigning each phoneme to the next available clockwise slot. . . . .* 101
- 5.1 *Three examples of stimulus lists from Treiman and Danis (1988, experiment 1) . . . . .* 121
- 5.2 *The proportion of responses showing various consonant-vowel structures in simulations of serial recall of lists of six monosyllabic nonwords.* 121
- 5.3 *Distribution of error types in simulations of serial recall of lists of six monosyllabic nonwords. The scheme used is similar to that of Bisiacchi, Cipolotti, and Denes (1989) . . . . .* 122
- 5.4 *The proportion of errors with different phonemic lengths. Data from Treiman (in press) are shown for comparison . . . . .* 122



## LIST OF TABLES

---

- 5.5 *Responses categorized according to the scheme adopted by Treiman and Danis, 1988 (T&D). 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.* 124
- 5.6 *The distribution of error types as a proportion of all responses, in simulations of nonword repetition. For comparison, the repetition errors made by four year old children (Gathercole, Willis, Baddeley, & Emslie, in press), and patients RR (Bisiacchi et al., 1989) and IGR (Caramazza, Miceli, & Villa, 1986) are shown. . . . .* 130
- 5.7 *Table showing the percentage of correct responses in simulations of nonword repetition for stimuli of different lengths. Mean data from children of ages 4-9 years (Gathercole, personal communication) . . .* 130
- 5.8 *The distribution of contextual and non-contextual substitution errors in the model's attempts to recall the first and last items in lists of 6 CVC nonwords (analysis includes only CVC errors). . . . .* 133
- A.1 *A table showing the form of local coding used for phonological information in the simulations reported in chapter 5. Also shown are the template slots that are activated (during learning) by each 'phoneme'. 156*
- D.1 *Table showing activations of template nodes during presentation of an unfamiliar phonological sequence . . . . .* 172

# 1. Introduction

---

## 1 General introduction

This thesis takes a computational approach to the problem of developing useful theories of short-term memory. Informally, theories of verbal short-term memory frequently suggest that it employs representations and processes common to the speech production system, yet formal models of short-term memory have never had much in common with models of speech production, which are generally shaped by psycholinguistic insights. Meanwhile, models of speech production have largely ignored the issue of how the linguistic structures they employ might be learned. The aims of this thesis are to develop an account of how, at the level of phonology, unfamiliar stimuli can be learned accurately in a single presentation; how linguistic constraints play a role in learning and recall and how such a linguistically-constrained model of verbal short-term memory is compatible with existing accounts of speech production and short-term memory. This general introduction explains why this direction has been taken, and fleshes out the philosophical starting points of the

project.

Briefly, the research described in this thesis has been guided by two important convictions regarding theory:

1. That concrete, testable theories are to be preferred over more abstract ones. Hence the decision to adopt a computational approach to modelling. To be successful, a concrete model should provide a detailed explanatory account of some hitherto unexplained phenomena, and if possible, generate novel and testable predictions.
2. That, where possible, theories should integrate the theoretical insights and basic empirical facts from different areas to produce workable models with ever-greater scope, rather than attempting to explore the finer points of the empirical data in an isolated domain. The work described in this thesis is an attempt to bring together theoretical and empirical work from three areas, in very general terms:
  - (a) Short-term memory.
  - (b) Serial order.
  - (c) The linguistic properties of speech.

## 2 Short-term memory

Psychologists often draw a distinction between tasks involving short- and long-term memory. Short-term memory tasks are those involving the rapid learning of unfamiliar stimuli, such as a randomly ordered list of digits, or a new word. While short-term learning is very rapid (e.g., a short list of numbers can be learned in

## 2. SHORT-TERM MEMORY

---

a single presentation), retention is temporary, such that any delay between presentation and recall results in a reduction in performance. In contrast, long-term learning takes much longer but can result in the retention of information for decades (for example, an adult can recall the names of childhood friends). Another important difference between long- and short-term memory is that short-term memory is subject to capacity limitations, no such limit is apparent for long-term memory.

The clearest demonstrations of the capacity limitations of short-term memory come from experiments which measure memory span. Such experiments typically involve the learning of short sequences of familiar items such as digits or letters. After a single presentation, it is the subject's task to recall the items in order. As the number of items in a stimulus list increases, the probability of correct serial recall is reduced. There is a sigmoidal relationship between list length and the proportion of attempts which are successful (Guildford & Dallenbach, 1925). Span is normally defined as the length of sequence which can be recalled accurately in 50% of attempts. For stimuli such as digit or letter sequences, span is around seven items in the normal population (Miller, 1956).

The most influential models of memory have been those which suggest that the distinction between short- and long-term memory performance reflects an underlying duality in the mechanisms of memory (e.g., Atkinson & Shiffrin, 1968), and although such models have not gone unchallenged (e.g., Melton, 1963), there is now a consensus in favour of models which in one way or another distinguish between subsystems responsible for permanent and temporary storage. Some of the strongest evidence for dual-store models comes from neuropsychological case studies concerning patients who show specific deficits of either short- or long-term memory performance (e.g., Baddeley & Warrington, 1970; Shallice & Warrington, 1970; Basso, Spinnler, Vallar, & Zanobio, 1982). Recently components of a specialised verbal short-term memory system have been localized in experiments using Positron Emission Tomography

## 2. SHORT-TERM MEMORY

---

(Paulesu, Frith, & Frackowiak, 1993). However, to regard short- and long-term stores as entirely independent parts of a modular system is an oversimplification. In almost all experimental tasks the possibility exists for short- and long-term subsystems to interact in producing a response. For example, memory span is greater for familiar words (for which long-term representations have been established) than for nonwords (Hulme, Maughan, & Brown, 1991).

Here we shall be concerned in the main with short term memory, and this being the case I shall concentrate on studies which so far as possible, isolate short-term storage by using stimuli which are entirely novel (such as nonwords). However, even here some long-term knowledge will be useful. In recalling a temporarily-stored nonword (e.g., “darliamentary”) the existence of a similar familiar word can aid recall (Gathercole, Willis, Emslie, & Baddeley, 1991). More central to the current work is the idea that existing phonological knowledge (knowledge concerning the linguistic structure of syllables) is used in the representation of new words. The nature of the relationship between long- and short-term storage, and the extent of any overlap between them remains somewhat unclear, and it will be one of the issues addressed in this thesis.

### 2.1 Functions of short-term memory

The fact that humans have a limited ability to learn and recall meaningless sequences of digits might be thought to be of little consequence, were it not the case that measures of span correlate positively with measures of performance on many less esoteric skills, such as reasoning, and comprehension. For this reason, tests of serial recall of familiar verbal materials, such as letters, digits and words, are frequently used in the assessment of general cognitive function. The implication is that performance on the serial recall task reflects the capacity of a store of limited capacity that is in

## 2. SHORT-TERM MEMORY

---

some way involved in a range of cognitive functions – a *working memory*. Because of its apparently central role in general cognition, working memory has received a great deal of attention in recent years (see Baddeley, 1986 for a detailed account of work in this field).

Less attention has been given to the nature of the relationship between long- and short-term memory. As well as playing an important role in complex tasks such as reasoning, verbal short-term memory is believed to be crucial in long-term learning. Recently experimental evidence of the link between short-term memory and vocabulary acquisition has emerged from a longitudinal study (Gathercole & Baddeley, 1989). It was found that children showing an early proficiency in nonword repetition (a very basic test of verbal short-term memory) subsequently acquired vocabulary more rapidly than their contemporaries. More evidence for the dependence of long-term learning on short-term storage comes from a neuropsychological case study which showed that in one patient with a specific short-term memory deficit, long-term learning of unfamiliar-familiar word paired associates (a task analogous to vocabulary acquisition) was extremely impaired relative to a similar task involving only familiar words (Baddeley, Papagno, & Vallar, 1988). In short, it appears that in order to learn a new word (or develop an association between a new word and a familiar one), a short-term representation of that word is necessary. Informal theories of memory have often suggested that long-term learning (for example vocabulary acquisition) depends upon the transfer of information from short- to long-term storage through some form of rehearsal or elaborative processing (Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972). There is some evidence to support this contention. For example, nonsense syllables are learned better if they are rehearsed by overt articulation (Mechanic, 1964). Long-term learning through rehearsal clearly requires a short-term representation which can support articulation.

A possible explanation of the functional role of the relationship between short- and

## 2. SHORT-TERM MEMORY

---

long-term phonological memory comes from a consideration of the computational problem that the learning of new words presents. In acquiring vocabulary, the novelty of the stimulus must first be detected, otherwise useful representations of known words will be needlessly reproduced; for a temporally-distributed input (such as a spoken word) this necessitates its temporary storage – it cannot be clear at the outset whether the word is novel or familiar, this can only be determined with reference to its context. For example, the sound “ank” in an utterance may represent a part of a familiar word in one context (“... weigh anchor, Mr. Christian ...”), or the whole of a novel one in another (“... the ancient egyptian ankh symbol ...”), and part of a novel phonological form in another (“... his knees were affected by ankylosis ...”); some consideration of the context in which the sound appears is required in order to determine whether long-term learning is necessary or not. However once novelty in the input has been detected, the stimulus has passed, therefore its temporal structure (or the aspects of it which are required to convey its meaning) must be “buffered” while long-term learning takes place.

Since the ability to repeat or rehearse new words emerges so early in development and appears to play such an important part in vocabulary acquisition, it would seem sensible to regard this temporary buffering of unfamiliar materials as one of the central functions of the verbal short-term memory system. The temporary storage of verbal materials necessary for vocabulary learning can later be exploited to support less basic activities such as reading and comprehension. It is thus important that any useful model of short-term memory be capable of representing new words (or nonwords) in a single exposure. A detailed model of this process is expected to shed some light on the way in which short-term memory is involved in long-term learning, and the way in which short- and long-term memory systems overlap.

### 2.2 The role of speech in short-term memory

There is strong evidence for the involvement of speech-related processes in short-term memory. For example, the relationship between span and articulation rate is linear (Baddeley, 1986):

$$S = c + kR \quad (1)$$

where  $S$  is span,  $R$  is articulation rate, and  $c$  and  $k$  are constants. Experiments have shown that span is dependent not simply upon the number of items in a sequence but on their spoken duration. For familiar materials span is approximately equal to the number of items that can be articulated in between one and two seconds (Vallar & Baddley, 1982).

For the brief storage of verbal materials, it appears that a speech-based form of coding is generally used: Evidence for this comes from the effect of phonological similarity on serial recall performance. Conrad and Hull (1964) demonstrated that recall for lists of visually-presented consonants was better for lists comprised of items that were phonologically dissimilar (e.g. "Q, R, N, W, H") than for items that were phonologically similar (e.g. "D, P, B, E, V"). The principle effect of phonological similarity is to increase the frequency of order errors (Wickelgren, 1965). If meaningful words (rather than letters) are used, then the semantic similarity of the items has comparatively little effect (Baddeley, 1966).

Speech processes also appear to play a role in maintaining a stable trace in short-term memory. The most important of these processes is believed to be rehearsal, an overt or covert repetition of the stimulus material which can be used to refresh its trace in short-term memory. Baddeley and Hitch (1974) suggested that verbal short-term memory could be fractionated into two components, a *phonological store* which held



## 2. SHORT-TERM MEMORY

---

a decaying trace of the stimulus, and an *articulatory loop* – a speech-based control process which serially refreshed the trace through rehearsal. The involvement of this articulatory process which is responsible for the relationship between span and articulation rate (equation 1). The model also helps to explain the finding that when rehearsal is prevented by a concurrent articulatory task (e.g. saying “the, the, the ...”), recall is adversely affected (Murray, 1967).

The detailed nature of phonological coding is a central issue in the present work. I shall argue that error data from short term memory studies suggest a form of phonological representation which is hierarchically-organized and constrained by linguistic principles. In order to generate a spoken response in a short-term memory task, it is obviously necessary to move from an static internal representation of the intended utterance to a serial pattern of action in the effectors. It is argued in this thesis that the internal representations are such that this can only be achieved by articulatory processing. A more detailed account of phonological coding may thus help to explain why it is that articulatory control is involved in verbal short-term memory.

### 2.3 Serial order and short-term memory

The need to rapidly develop representations of novel spoken stimuli presents a problem: speech stimuli, unlike say static visual stimuli, are temporally-distributed. At no time is the information representing the whole utterance available concurrently. When the utterance is finished, the stimulus is gone, and cannot be re-examined. Yet, despite the fleeting nature of speech stimuli, normal subjects, even children as young as two and three years old, can often repeat unfamiliar words after a single presentation (Gathercole et al., in press). In so doing, they demonstrate the remarkable capacity to form an internal representation of the serially-ordered stim-

ulus 'on-line', as the stimulus itself occurs.

It is therefore essential that any implemented model of verbal short-term memory consider the question of how serial order is represented mentally, and in particular how it is represented. The storage of serial order information is an important property of short-term memory, and this is reflected by the use of memory span as an measure of short-term memory performance. It is a sensitive test because the majority of errors in this paradigm are order errors (Bjork & Healy, 1974), that is errors where the identity of part of the intended sequence is preserved, but its serial position in the output sequence is wrong. In fact, across a range of tasks in which serial order is critical, humans are apt to make ordering errors. The task of learning new words involves learning a sequence of speech sounds rather than a list of digits or letters, but computationally, the problem appears initially to be very similar. Thus a consideration of some general models of serial order is essential in developing a model of verbal short-term memory.

### 2.4 Errors in speech and short-term memory

The way in which a system breaks down can help to shed light on its internal structure. To take an example from outside psychology, imagine being in a restaurant in which some music is playing. It is impossible, from your table, to observe the source of the music, because it is in another room. You may entertain a number of hypotheses about its nature; it could be a radio, a record-player, a compact-disc player, a tape-recorder, even a band. While the music is playing normally, it is difficult to infer much about the nature of the system which is responsible. If the system breaks down however, it may be possible to make some inferences about the system, because the different hypothetical systems are each prone to different characteristic faults. If, for example, the same short fragment of music is played over and over

## 2. SHORT-TERM MEMORY

---

again, it suggests that some there is something cyclical about the way the music is represented in the system, and that the same cycle is being accessed in error over and over again.

A similar problem faces the psychologist. We are hampered in our attempts to understand the mechanisms of cognition because, despite recent technological advances, the internal structure of the functioning brain remains, to a large extent, inaccessible. We have contact with human cognitive systems through peoples' behaviour. We can entertain a number of models about a given system. Some may provide a better account of the basic competence of the system than others, and we may favour some because they are more elegant or parsimonious. It will, however, be difficult to distinguish between competing theories which adequately describe the correct operation of the system. The character of the errors to which a particular hypothetical system is prone can be compared with those made by experimental subjects, and thus error data can be brought to bear on the theoretical argument.

The form that ordering errors take can help us choose between different models of serial order. Chapter 3 discusses experiments which show how people make mistakes in producing sequences of speech sounds. These errors are different to those found in recalling other kinds of sequence (for example, digit lists). However, there is a good deal of similarity between patterns of error in the recall of nonwords and in the spontaneous production of speech. Both types of error are constrained by linguistic principles governing syllable structure. It may thus be fruitful to consider speech production as being a special case of serial recall, where linguistic constraints affect the manner in which sounds are sequenced. Ideally, one would like to develop an account of verbal short-term memory which while compatible with at least one general accounts of serial order, is capable of dealing with the special factors which apply to speech.

## 3 Overview

The next chapter shows how many existing accounts of serial order are flawed by their inability to account for typical ordering errors. Other models provide a more promising approach, but cannot account for the pattern of constraints found in phonological ordering errors. As outlined above, these constraints are likely to provide important clues about the underlying structure of the system responsible for serial order in speech and short-term memory. Error data from spontaneous speech and from short-term memory experiments are compared and discussed in chapter 3. These data provide the basis for the constrained articulatory encoding and output system which is developed in chapter 4. This system will be incorporated into the new model described in chapter 5 which also draws upon insights from some of the models of serial order discussed in chapter 2. In chapter 6 further issues arising out of the new model are discussed, including the light it sheds on the nature of the relationship between short- and long-term memory.

## 2. The problem of serial order: general issues and models

---

Experiments devised to investigate short-term memory have often focussed on the difficulty in maintaining a representation of unfamiliar sequences. In serial recall for example, the most common type of error is a misordering of the items in the target list. It is the novelty of the sequence, rather than its content, which seems to test short-term memory. If a well-known sequence such as the months of the year is presented, it is very likely to be recalled correctly. If the same items are presented in random order, the probability of correct recall is clearly greatly reduced.

Though the representation of unfamiliar sequences seems to pose a particular problem, ordering errors are not restricted to tasks involving retrieval from short-term memory. Even where the retrieved material has a long-term representation, order errors account for a large proportion of observed errors. Thus in typing or spelling familiar words many errors involve the transposition of letters (see e.g., Shaffer, 1976; Caramazza, Miceli, Villa, & Romani, 1987). Spontaneous speech errors often

involve the misordering of small phonological segments of the intended utterance (e.g., Shattuck-Hufnagel, 1979).

A capacity for complex temporally-organized behaviour is one which distinguishes humans and other higher animals from simpler organisms (Lashley, 1951). Linguistic communication and reasoning are examples of serial behaviours which are often regarded as quintessentially human characteristics. Yet, despite our undoubted prowess in processing serial information our performance is still characterised by ordering errors, particularly when we try to learn unfamiliar sequences.

Since the representation of serial order is such a general problem, a general theoretical framework may help to explain the basis of ordering errors in a range of tasks. On the other hand, some aspects of the error data in any particular task may reflect task-specific constraints. In this chapter, the aim is to identify the important theoretical issues which the general problem of representing serial order throws up. Constraints on serial order specific to the representation of speech will be dealt with in chapter 4.

The problem of serial order has not always received the theoretical attention which it deserves; in recent years there has been a tendency for models to make vague reference to 'buffers', systems dedicated to the storage of often serially-ordered information, without further elaboration of their internal structures (such models are considered briefly below). However, there have been attempts to advance general psychological theories of serial order (e.g., Lashley, 1951; Wickelgren, 1969, 1976; Shaffer, 1976; Estes, 1972; Grossberg, 1978; Amit, Sagi, & Usher, 1990; Houghton, 1990). In addition several models have been put forward which develop the general theories in the context of specific empirical domains (e.g., typing, serial recall, spelling and speech production: Rumelhart & Norman, 1982; Burgess & Hitch, 1992; Houghton, Glasspool, & Shallice, 1994; Dell, Juliano, & Govindjee, 1993).

## 1. 'BUFFERS'

---

This chapter discusses some of these models, and the issues that are thrown up by serious theoretical consideration of the representation and control of serial order. In almost all of the models ordering information has been represented in one of two distinct ways; either in terms of the temporal relations *between items* in a series (which I will refer to as item-item or I-I models), or in terms of an external representation of the serial position or temporal context of each item (position-item or P-I models).

### 1 'Buffers'

In computing, a buffer is a structure which is loaded with the elements of a stored series. The structure itself specifies the serial order in which the stored information will be output. The control process which accesses a buffer simply steps from 'left to right', activating each stored element in turn. Serial processes like this are provided by programming languages. In artificial intelligence and computing, serially-ordered structures such as buffers are available as primitives.

The growth of cognitive psychology from the early sixties to the present has been accompanied by a movement towards theories based on the computer metaphor for the mind. One unfortunate consequence has been the credibility given to psychological theories which manage to avoid the issue of how serial order is represented mentally by appealing to technological devices. Many theories make reference to 'buffers', used to store (often serially-organized) information as it is being processed, while making no attempt to further elaborate their internal structure/mechanism. Although it may well be the case that such these psychological 'buffers' play a similar functional role to their technological counterparts, there is little reason to suppose they will have much in common at a mechanistic level. The buffers of classical digital architectures are not really plausible candidates as models of the psychological

## 1. 'BUFFERS'

---

mechanisms supporting serial order, and they have never been seriously defended or tested on empirical grounds. Instead they merely serve to obscure the inadequacy of many theories which are not specified to the extent that they can explain the occurrence and distribution of ordering errors.

Where error data are addressed it soon becomes apparent that the use of a buffer metaphor is rather unhelpful, and it becomes necessary to elaborate its internal structure and to introduce mechanisms which differ substantially from technological ones. Although progress is likely to be hindered by the central inadequacy of the computing metaphor for serial processing, this process can lead to useful insights. Shaffer (1976), for example, proposed a model of typing that made use of 'buffers'. Yet, in order to account for error and response latency data, he found it necessary to postulate the existence of serial processing mechanisms somewhat removed from those used in computing. According to this model for each word to be typed, there is a translation operation which converts a symbol representing the entire word (in an input buffer) into a list of symbols standing for keypresses in the output buffer. The order of the listed symbols is determined by an index associated with each which designates its serial position. The symbols are copied into the output buffer in the specified order.

Shaffer accounts for typing errors in terms of four different types of malfunction; translation errors, execution errors, pointer errors, and indexing errors. From the point of view of the work described in this thesis, indexing errors are the most interesting category: in these errors it is assumed that the indices of one or more letter tokens are corrupted (in some undefined way) such they become confused with other similar indices (through some other undefined mechanism). Each index represents not just the ordinal position of the letter, but also its position relative to syllable boundaries. This was necessary to account for a constraint on typing errors: when letters move from one syllable, they take up an analogous position within



another (e.g., *forget me knots* → *forget me fnots*). This suggests that the indices of letters occupying similar syllable positions in different syllables are confusable. The index mechanism is never specified in any detail. However, Shaffer clearly recognises that the representation of serial position within the output buffer must capture linguistic properties if the error data are to be understood. The use of a classical symbol processing architecture hinders further progress, since it is not clear how the suggested ordering mechanisms might be implemented.

The shortcomings of the buffer analogy become more apparent when one considers the other error categories. Translation errors occur when the contents of the input buffer are incorrectly translated into output symbols (a kind of catch-all category, since it would seem that this could result in any outcome). Execution errors result from failures at the motor level (e.g., missing a key, pressing two keys simultaneously etc.). Pointer errors occur when the position of the 'program pointer' has not been properly updated after the output of a symbol, so that it either overshoots the next position, or fails to move forward at all. The mechanism responsible for updating the pointer's position is not specified, so it is not easy to see why the pointer might fail in this fashion rather than, for example, visiting serial positions in random order. In each case, the failure of a different subsystem has been proposed to account for a type of error, but in each case the mechanism responsible for the error is not specified. Therefore, it is impossible to see how and why the subsystem fails in the way it is supposed to, or indeed how it operates normally.

## 2 Associative chaining

Some of the earliest psychological accounts of serial order (e.g., Ebbinghaus, 1964) postulated that action sequences were represented as unidirectional stimulus-response (S-R) links. The appeal of this type of account is its simplicity; it requires nothing

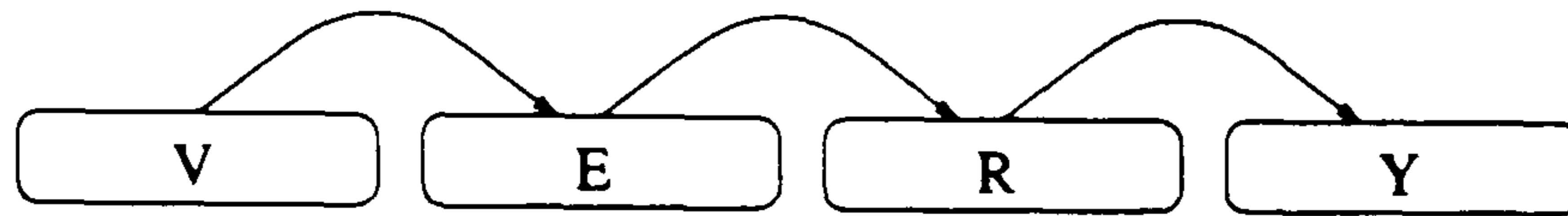


Figure 2.1: *Representation of a sequence in the form of a chain of item-item associations.*

more than a representation of the items themselves and the links between them. Retrieval of a sequence is achieved by tracing a path through the links. Figure 2.1 shows how a sequence of letters (“V, E, R, Y”) could be represented.

### 2.1 Representing repeated items and multiple sequences

In a wide-ranging criticism, Lashley (1951) identified a number of serious shortcomings in the chaining approach to serial order. One of the major difficulties concerned the representation of an item which occurs in more than one serial position. This problem arises in two contexts:

1. *Representing sequences containing repeated items.* In such sequences, a stimulus action is associated with more than one response, and chaining models provide no mechanism for choosing between different associative links. This is illustrated in figure 2.2a, which shows the associative links necessary to represent the sequence “E, V, E, R, Y”. “E” must be linked to both “V” and “R”, so that it is not clear which item follows the first instance of “E”. This problem arises again when the second instance of “E” is realised, and there is the potential for endless looping through the first part of the sequence, without ever reaching the final item.
2. *Representing multiple sequences in the same associative structure,* for example sequences of speech sounds making up the familiar words in a mental lexicon. If the same elements are to be used to stand for the /t/, /æ/, and /k/ in tack,

## 2. ASSOCIATIVE CHAINING

---

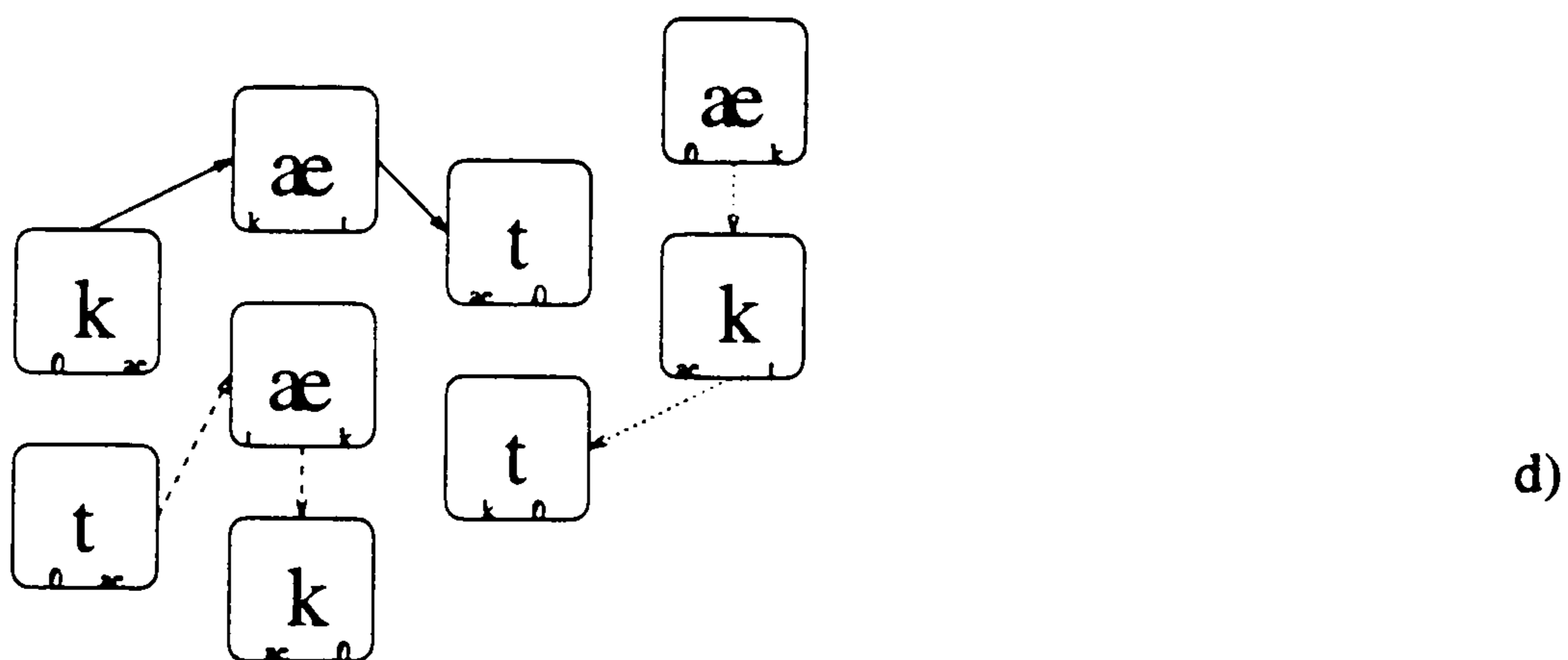
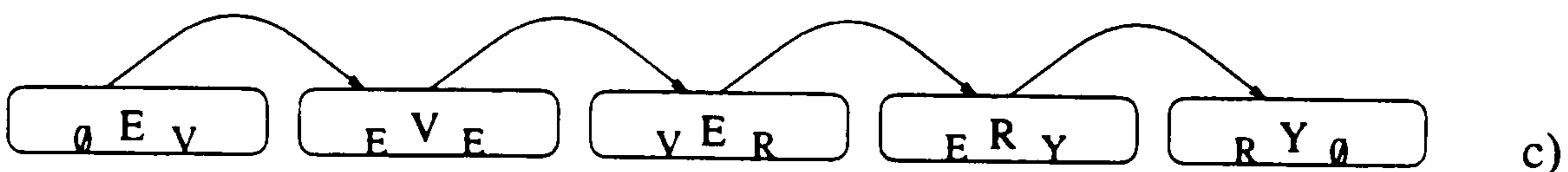
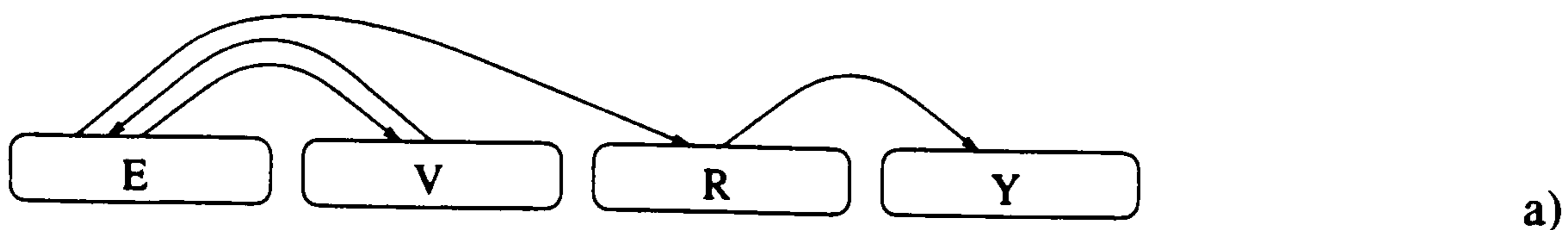


Figure 2.2: a) Associative chaining has difficulty in accounting for the representation of sequences containing repeats. It is not clear which of the associative links from “E” should be followed. b) If a single associative structure is to represent multiple sequences, there is no mechanism to select those links which must be followed to generate a particular sequence. Here phonemic representations of cat, tack and act would interfere with one another (different types of arrows are used to show the links needed for each word). c) and d) Wickelgren’s context-specific coding overcomes these difficulties, but only at the expense of using entirely different tokens to represent instances of the same type

## 2. ASSOCIATIVE CHAINING

---

cat and act, then the chains will be linked together to form a network (see figure 2.2b), in which the underlying serial structure of any single word is obscured by links between them.

Wickelgren (1969) suggested a form of context-sensitive chaining which avoids some of these difficulties. In this account, elementary actions are represented by different tokens depending on their immediate context. For example the sequence “E, V, E, R, Y” would be represented  $\emptyset E_V, E V_E, V E_R, E R_Y, R Y_\emptyset$ . Because Wickelgren’s approach to chaining uses *tokens* (e.g., standing for particular instances of each “E” in the example) rather than *types* (standing for the category of actions designated “E”), a chain involving a repeated action can be represented without linking one stimulus to more than one response. In figure 2.2c). It is clear that by following the stimulus-response chain from left to right, the target sequence “E, V, E, R, Y” can be generated. However, the token based form of representation is immediately unappealing, because it fails to capture any relationship between different instances of the same item in a sequence. In the above example, the two “E”s are as different from one another as they are from the other letters in the sequence. In fact there is no reason why the action represented by  $\emptyset E_V$  should resemble the action represented by  $V E_R$  in any way. The scheme thus does allow for different orderings of the ‘same’ actions within the same associative structure, but only by suggesting that the *same* actions have quite *different* internal representations.

Figure 2.2d shows how context-sensitive coding deals with the problem of representing multiple sequences in the same associative structure. Here, associative chains representing cat, tack, and act do not interact. This is because allophonic variants of each speech sound have quite different representations in Wickelgren’s scheme. For example, the /æ/ in “cat” is represented by a token designated  $_k \alpha_t$ , whereas the /æ/ in “tack” is represented by completely different token,  $_t \alpha_k$ . These units have been termed ‘wickelphones’. The use of different tokens for variants of the same phoneme

can be used to provide a weak account of coarticulation simply by assuming that each wickelphone is associated with a different articulatory realization. However, Wickelgren's account fails to provide any explanation for the *similarity* between the same phoneme occurring in different contexts (e.g., the /æ/ sounds in cat and tack from the above example).

In addition to the unsatisfactory use of a token-based representation, Wickelgren's solution to the problems chaining models face with repetition is incomplete. If an action is repeated with identical local context (e.g., in /kænkæn/), the same wickelphone must be used twice, creating the kind of looping chain shown in figure 2.2a. In addition it fails to solve the related problem of representing multiple sequences in the same associative structure solved. For example, all sequences beginning /kæ.../ would begin with the same wickelphone  $\emptyset k_{\text{æ}}$ . Thus to generate the sequence "catalyst" it would be necessary to choose between associative chains radiating from the same starting point (/kæt/, /kæməl/, /kæneəri/ etc.).

### 2.2 Coarticulation and Anticipation

A further problem Lashley (1951) had identified in associative chaining models was their failure to account for phenomena which suggest that elements are active at some level before they are executed. These phenomena can be observed in a range of serial behaviours and fall into two categories:

1. *Coarticulation and analogous effects.* In many serial tasks, actions scheduled to be executed later in a sequence appear to exert an influence on realisation of earlier actions. In speech, where they are ubiquitous, the effects of this pre-emptory influence are called coarticulation. For example, in producing the word "stew", the lip-rounding required to produce the vowel sound precedes

the vowel and alters the form of the consonant cluster. Similar pre-emptory movements are observed in typing (Rumelhart & Norman, 1982).

Coarticulation in tasks requiring serial order may reflect a more general organisational requirement for response systems; potential actions must be represented in parallel so that responses can be ordered in situations where they are incompatible with one another and thus cannot be performed simultaneously. For example, one cannot simultaneously reach toward two separated targets with the same hand. Experiments have shown that in situations like this, the presence of distractors influences reach trajectory, indicating that there is parallelism in the representation of potential responses even when the task does not demand a sequential response (Tipper, Howard, & Jackson, submitted).

2. *Anticipatory errors.* The simple stimulus-response relationship between elements in an associative chain suggests that one element cannot become active before its predecessor, yet many errors in serial tasks involve the execution of an action in advance of its intended serial position (for example in typing, *full of heavy* → *fully of heavy* or in spontaneous speech “there has been some paint spent” → “there has been some spaint spent”). Even if it were possible for an element to be activated before its stimulus, one would expect the remainder of the chain to be followed as normal, resulting in an omission rather than a substitution (e.g., “There has been some spent”). Exchange errors (e.g., *pastoral* → *pastorla*) are even harder to account for: not only is one element anticipated, but the one it displaces is subsequently activated later in the sequence. As we shall see, these errors are much more comfortably dealt with by other models.

Taken together, these phenomena provide strong evidence of a degree of parallelism in serial output systems. In the simple S-R chaining models outlined above there is no sense in which items in a sequence are ‘active’ before they are executed.

### 2.3 More recent 'chaining' models

Despite the apparent validity of Lashley's widely-cited, forty year-old critique, variants on the theme of associative chaining remain amongst the most influential accounts of serial learning to the present day. The main development has been the elaboration of mechanisms for the *learning* of chaining associations, which were absent from early accounts. These learning mechanisms are taken from models of simple associative learning. By allowing that each response can serve as the stimulus for another, chains of association can be developed using the same machinery as is required for stimulus-response learning. For each instance of a theory describing the acquisition of simple S-R associations, there is a theory which implements serial order by using the learning mechanism to develop a chain of such links. Examples include Jordan's (1986) connectionist architecture (based on a three-layer perceptron, and using backpropagation learning), a similar architecture due to Elman (1990), Lewandowsky and Murdock's (1989) model, based on Murdock's (1982) Theory Of Distributive Associative Memory (TODAM), which uses vector convolution to develop item-item associations, and models based on Hopfield learning (Amit et al., 1990).

As well having an appealing generality (sequence learning is regarded as a form of S-R learning), these models retain the parsimony of all chaining models; there is no 'unnecessary' machinery – while I-I models require a only links between a set of items, the alternative P-I accounts would require a representation of serial position, as well as the items themselves and position-item associations. However, the mechanistic models retain most of the shortcomings of the chaining approach, as well as its appeal. To date, none has offered a convincing solution to the problem of interference between sequences, or to the difficulty in representing sequences containing repeated items. Neither do they predict the error types characteristic of human performance on serial tasks – item transpositions. These fundamental

flaws would appear to result from the form of item-item representation used in the models. Lashley's (1951) conclusion that associative chaining models were "doomed to failure" has not been seriously challenged. Progress has, however, been made through the development of models employing parallel processing. Some of these models are described below. They were developed, in part, to provide explanations for the coarticulatory phenomena described in section 2.2. Initially, they developed along the lines of chaining models, using item-item associations to represent serial order, with the inherent problems that this brings. More, recently however, models have emerged which avoid the pitfalls of I-I models, by using position-item representations of serial order instead.

### 3 Parallel models of serial order

Models employing discrete formalisms have not proved very successful in illuminating the basis of serial order errors; there is a good deal of data suggesting that actions to be sequenced are in some sense simultaneously active before they are executed. In the buffer theories and associative chaining models described above, an action is either active (being performed) or inactive. Connectionist models, by contrast, start from the premise that parallel processing is possible. The common feature of connectionist models is that cognitive processes are thought of as distributed over a (usually large) number of fairly primitive processing elements which interact with one another through a pattern of connections. If these elements were to represent items in a sequence, then the potential for parallel activation could provide a natural account of the priming effects (coarticulatory phenomena and anticipatory errors) which are so characteristic of serial behaviour.

Figure 2.3 shows a small part of a typical connectionist network. Associated with each node is an activation value, which represents its current state, a set of afferent



### 3. PARALLEL MODELS OF SERIAL ORDER

---

connections, through which it receives information from other nodes, or the environment, and a set of efferent connections, through which it disseminates information to other nodes in the network. The processing power of each unit is very limited: it computes an activation value, usually a monotonically increasing function of its net input. This activation value may be used to calculate an output value, or the activation value itself may be output. The efferent connections propagate the computed value to other nodes in the network. These connections are associated with a strength or weight which determines the degree to which the output value computed by a particular node contributes to the net input of each node to which it is connected. Learning can be effected by changing the weights systematically in response to input patterns. The learning rules, which control this process, typically make use of information which is available locally at each connection, such as the current activation values of both transmitting and receiving nodes, and often the prior strength of the connection. In many models, a teaching input is also utilised. This teaching input is computed externally (not within the network), and reflects the degree to which a node's response to its input differs from some 'desired' response.

Connectionist models of serial behaviour broadly follow the pattern laid down by earlier psychological models. They fall into three categories:

1. Models that avoid the problem of serial order. In connectionist accounts this is usually achieved by representing temporally-distributed information as if it were simultaneously available (spatially-distributed) during learning (and produced simultaneously during recall). Models adopting this approach are clearly unlikely to shed any light the empirical phenomena relating to serial order such as patterns of order error in serial recall, and they will not be discussed further.
2. I-I models which in one way or another represent the temporal relations between items, and thus their serial order. More recent accounts show how such

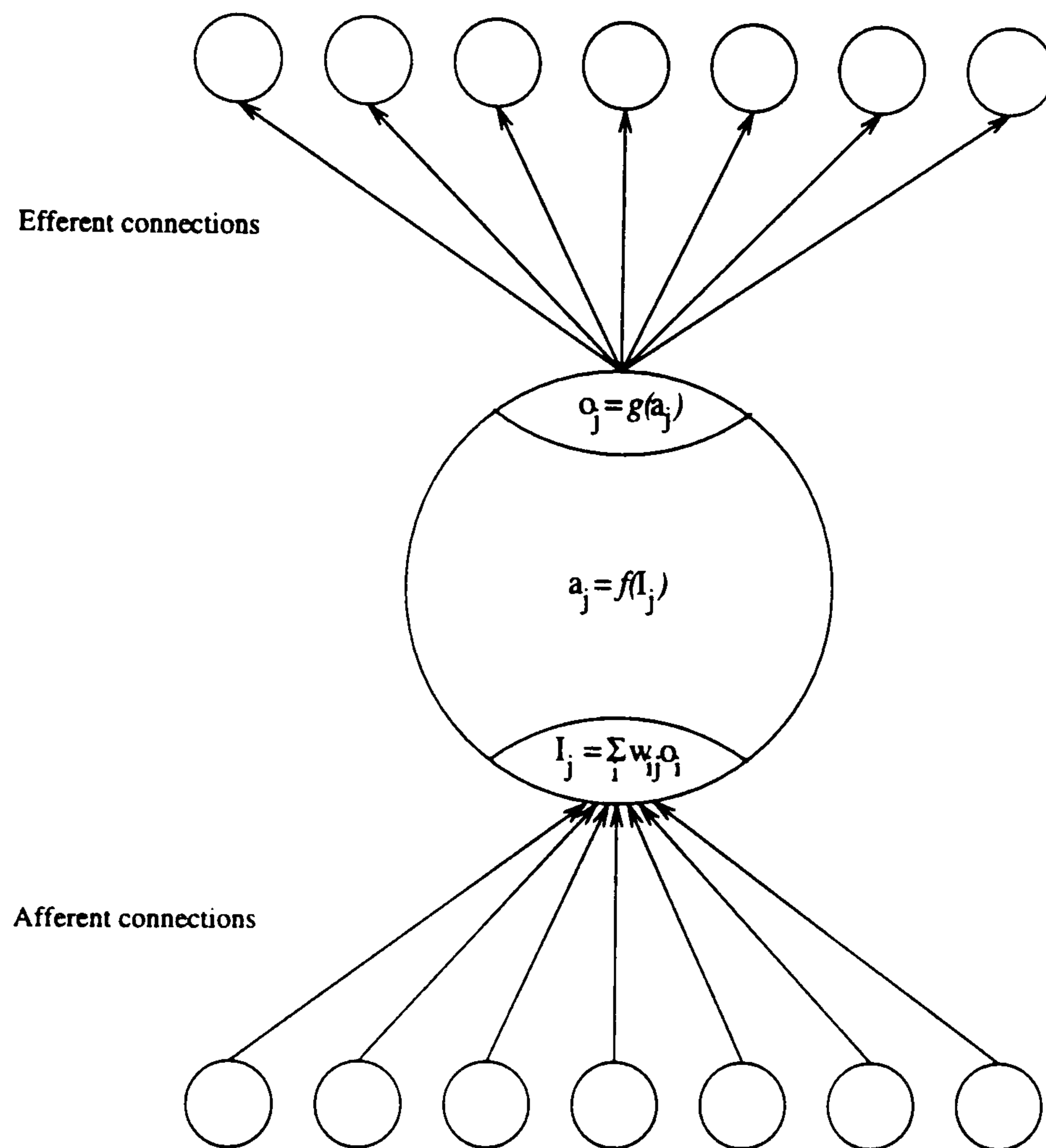


Figure 2.3: *Structure of part of a typical neural network. A node  $j$  has a number of afferent connections. The weighted sum of these inputs ( $i_j = \sum_i w_{ij} o_i$ ) is used to determine  $j$ 's activation ( $a_j = f(I_j)$ ). The node's output is a function ( $g$ ) of its activation. Each node's efferent connections are weighted, and weights can be changed systematically so that the network learns to produce an adaptive response to its inputs.*

associations can be developed through exposure to sequential input, using variants of well-known learning algorithms. In general, this has been achieved by making available to the learning rule, information about prior states of the network.

3. P-I models which specify structures and mechanisms for the representation of serial position, for position-item associations, and for the control of serial behaviour.

#### 3.1 I-I models

Estes (1972, see figure 2.4) suggested one of the first connectionist schemes for the representation of serial order. This general approach was later implemented by Rumelhart and Norman (1982) in their model of typing. The model used a local form of representation, with single nodes standing for both actions to be sequenced (keypresses), and sequences themselves (familiar words). All of the nodes representing actions in a particular familiar series of keypresses are activated in parallel by uniform connections from a *single* node standing for that sequence. An activation gradient is established by means of lateral inhibition which is arranged such that item nodes inhibit all other item nodes representing actions later in the sequence. The effect of the 'top-down' activation from the sequence unit is most attenuated in nodes representing items to be produced towards the end of the sequence.

Each item node tends to drive the output system towards a particular goal state (pressing the corresponding key). The more active nodes are more influential, and the output system thus moves towards the state associated with the most active item. When it reaches a state sufficiently close to one of the target states, the node representing that state is strongly suppressed. It then ceases to inhibit subsequent responses, allowing them to take control of the output system. Although the most

### 3. PARALLEL MODELS OF SERIAL ORDER

---

active node dominates the output system, at any time more than one item node can be active, and thus influence the state of the output system. Because of this, the network exhibits the kind of pre-emptive movements which are typical of human-performance in well-learned sequential motor tasks. A similar account could be given of coarticulation in speech, with the priming of forthcoming speech sounds tending to drive articulators in advance of their actual execution. Since an activation gradient is used to determine the order in which actions are selected for output, anticipatory errors might also be explained by assuming a degree of noise in the activations of the item nodes. This will occasionally lead to an item being selected in advance of its intended position.

The parallel activation of a set of elements representing the items in a sequence is essential to the representation of serial order in the Estes' model. This parallelism also helps to explain the priming and coarticulatory effects that in part prompted Lashley's (1951) rejection of associative chaining models of serial order.

However, like earlier models discussed above, the Estes account still relies on item-item associations to represent serial order, though in this case they are inhibitory. As a result it retains some of the problems of the simple chaining models. In particular the approach is not suited to the representation of sequences containing repeated items, where the lateral connections required for one part of the list conflict with those needed for other parts of the list, for example in the sequence "E, V, E, R, Y", "E" must be inhibited by "V" for instance, yet "E" is supposed to be the most active element at the outset. To overcome this limitation in their *implemented* model of typing, Rumelhart and Norman (1982) found it necessary to include a parsing mechanism which split words into parts in which no letter was repeated. The representation of many different sequences within the same associative structure is also difficult; as in the simple associative chain, the inter-item connections necessary for the representation of one sequence will often interfere with those required for another containing some of the same elements. These difficulties can, of course, be

### 3. PARALLEL MODELS OF SERIAL ORDER

---

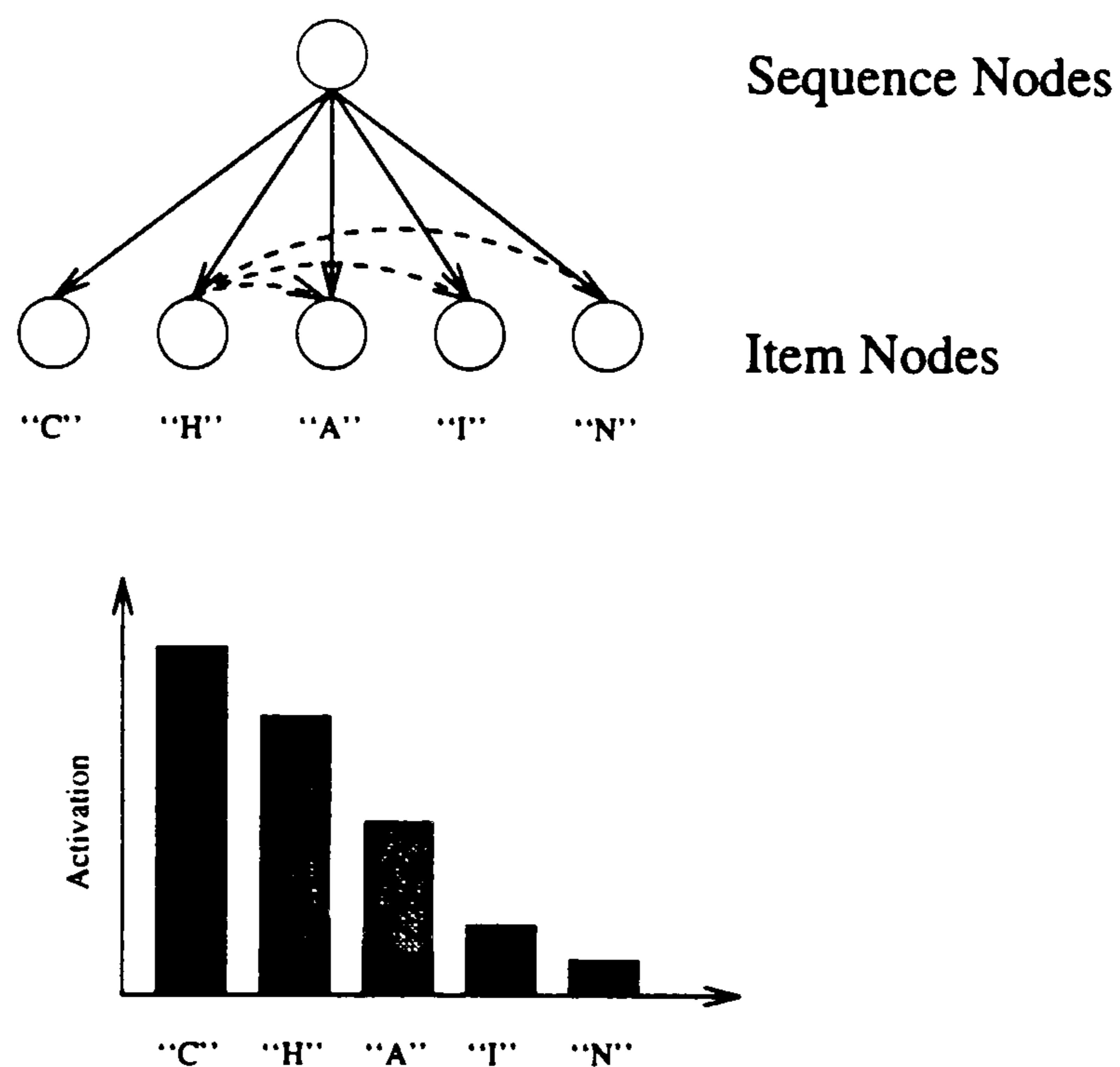


Figure 2.4: *Estes' (1972) scheme for the representation of serially ordered actions. For each sequence (e.g., "C,H,A,I,N"), there is a node which activates its nodes standing for its constituent items via excitatory links to each (solid arrows). Item nodes representing actions later in the sequence receive inhibitory input (dashed arrows) from nodes representing earlier items – for clarity, only the lateral connections from "H" are shown here; the later an item is in the list, the more items precede it. Items nearer the end of the sequence receive most inhibition, thus when the item nodes are activated in parallel by input from the sequence nodes, an activation gradient is set up.*

### 3. PARALLEL MODELS OF SERIAL ORDER

---

solved by proposing that tokens, rather than types are used to stand for the items in the sequence (c.f. Wickelgren, 1969). However, as Jordan (1986) points out, this will likely result in the combined activations of different instances of the same type overwhelming the output system. When producing the sequence "A, B, B, B", for example, the multiple instances of "B", all partially active, are likely to be sufficient to take control of the output system, producing "B" in the initial position rather than "A".

Another difficulty with the Estes/Rumelhart and Norman model is that it does not explain how the sequence specific lateral connections used to encode serial order might be learned. One approach to serial learning has been to adapt architectures originally conceived for the learning of static stimulus-response associations. One of the most versatile of these is the three-layer perceptron, and variations on this theme due to Jordan (1986), and Elman (1990) are probably the most widely-known connectionist accounts of serial learning.

The Jordan and Elman architectures are similar, and are illustrated schematically in figure 2.5. The essential difference between the serial networks and earlier models used to describe non-serial processes is the addition of *recurrent* or *feedback* connections, which allow information about prior states of the network to be made available to its input.

In the Jordan architecture (figure 2.5a), input patterns are composed of two components, plan and state information. To learn a sequence, the activation pattern of the plan units is held constant. The pattern of activation across the input nodes is fed forward, producing a pattern of activation in the output nodes. The pattern of activation in the output nodes is transmitted back to the state nodes via one-to-one fixed-weight connections. In addition, each state node feeds back to itself, so that the activation pattern of the state nodes is an exponentially decaying trace of

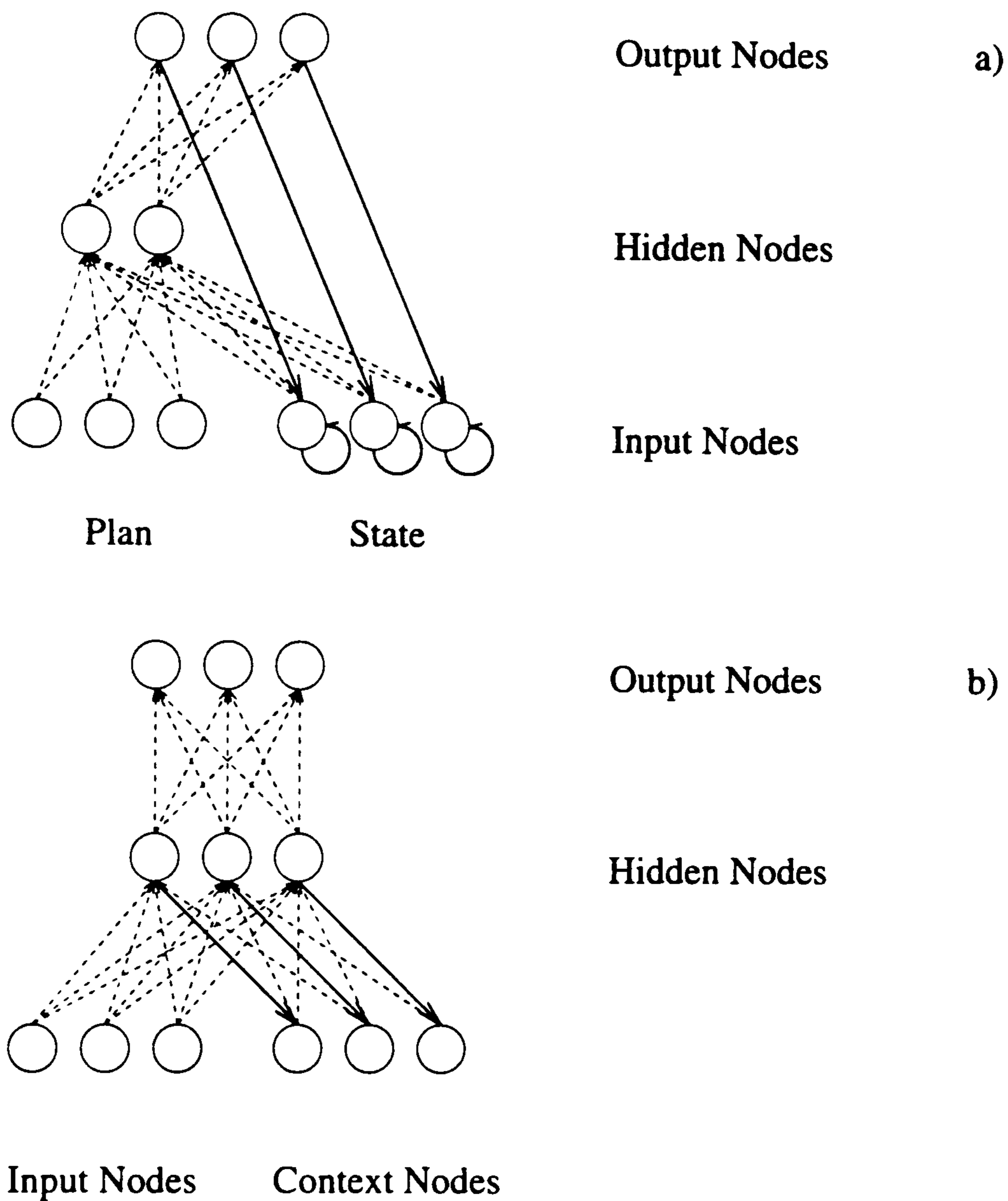


Figure 2.5: *Jordan (1986) and Elman (1990) have put forward similar schemes for the learning and representation of serial order in connectionist networks. In each case fixed recurrent connections (solid arrows) provide a time-varying context at the level of the input units. Modifiable feedforward connections (dashed arrows) are adjusted using supervised learning algorithms.*

### 3. PARALLEL MODELS OF SERIAL ORDER

---

previous output patterns. Successive inputs to the network thus comprise a static component representing the entire sequence (the plan), and a dynamic component representing prior outputs (the state or context). The dynamic component drives the network through a sequence of outputs.

A supervised learning algorithm is used to adjust the feedforward weights as each actual output is compared to a target output. In order to do this feedback from the output units is essential. Without it, the fixed pattern of activation in the input units would produce a fixed pattern of activation in the output units. The feedback connections allow the network to modify its response to the plan depending upon prior outputs, represented in the state nodes. Feedforward connections from these units will develop connections via the hidden units that tend to produce the target output which immediately follows the current output, that is forward chaining associations. At first, output states will bear no resemblance to the target outputs. Following repeated presentations, the network will produce the desired sequence of outputs whenever the appropriate plan is activated.

The Jordan architecture implements a form of associative chaining, as the recurrent connections between output and input layers allow one response (output pattern) to serve as part of the stimulus for the next. Elman's (1990) approach (see figure 2.5) is similar; although the direct association between stimulus and response is not as clear, it is nonetheless true that each input pattern determines the next. Instead of basing each response on the last (feedback from the output to the input units), the recurrent connections in Elman networks copy the state of the *hidden units* back to the input layer, to provide the dynamic context required to drive serial behaviour. The hidden units represent neither stimuli nor responses but an intermediate stage in processing. They thus come to represent task-dependent structural regularities in the input sequences. This means that domain-specific properties can be demonstrated (Elman, 1990).



### 3. PARALLEL MODELS OF SERIAL ORDER

---

As general theories of serial learning, the Jordan and Elman architectures have a number of serious problems. Perhaps the most important from the point of view of the current work, is that they are incapable of single trial serial learning. Both require multiple exposures to a stimulus sequence to learn it at all. The iterative learning algorithm is particularly important for the learning of sequences containing repeated items, which pose problems for both architectures. Consider a sequence like "E, V, E, R, Y", where "E", "V", "R" and "Y" correspond to different input patterns. The problem for the Jordan architecture is that the network must produce V in response to the first instance of "E", and "R" in response to the second. Since "V" and "R" are different, each instance of "E" must produce a different pattern of activation in the context nodes. However, the output patterns corresponding to both instances of E are the same. This is really an illustration of the difficulty in choosing between two potential successors in an associative chain. In the Jordan network context states cannot simply be copies of the most recent output pattern, otherwise responses following different instances of the same item will inevitably be the same. This is why the context units must feed back on themselves. The context state is not affected solely by the last output, but by all previous outputs. Note though that there will always be a strong positive correlation between such context states. This means that the network's response to the first instance of "E" will tend to interfere with the the response to the second instance (and *vice versa*). To learn to produce different outputs following each, the weights in the network must be adjusted to counteract this tendency. Thus, learning sequences containing repeated items will takes longer than learning sequences which do not (Jordan, 1986).

The Elman network offers a different, perhaps more elegant, solution to the problem. Here, the state of the context is determined, not by the previous output of the network, but by the previous state of the hidden units. There is no necessity for hidden units to have the same state every time a particular pattern (e.g., "E") is to be produced, so the network can develop connections which tend to produce

### 3. PARALLEL MODELS OF SERIAL ORDER

---

unique patterns of activation in the hidden units, for the same input occurring in different temporal contexts (following Wickelgren, these might be denoted  ${}_0E_V$  and  ${}_VE_R$ ). However, the solution of this problem requires many presentations of the input sequence.

As general theories of serial learning, these connectionist 'chaining' models are undermined because they are incapable of the single-trial, unsupervised learning which humans demonstrate in short-term memory tasks such as serial recall. While there have been attempts to develop connectionist implementations of chaining which use one-shot, unsupervised learning (e.g., Amit et al., 1990) these have not been able to overcome the central difficulty chaining models face when required to produce the same item in more than one serial position.

Because the models described in this section are capable of parallel processing, they avoid some of the pitfalls encountered by earlier discrete models. Effects such as coarticulation can be readily understood in terms of parallel activation (Rumelhart & Norman, 1982). The use of an activation gradient to represent the order of forthcoming actions could also help to explain anticipatory errors. There remain though numerous problems with these accounts, most of which arise because of their reliance on item-item associations to represent serial order. In particular, none of the models provides a very robust method for handling sequences containing repeated items. Those architectures which can deal with repeated items can do so only by exploiting the power of biologically-implausible learning algorithms such as backpropagation (e.g., Jordan, 1986), which are incapable of single-trial learning.

The weak points of the I-I models come, in the main, from their reliance on a sequence-specific representation of serial order, which tends to produce interference within and between sequences. Typical ordering errors are not at all natural to these models. Despite their appealing parsimony they are fundamentally incompatible

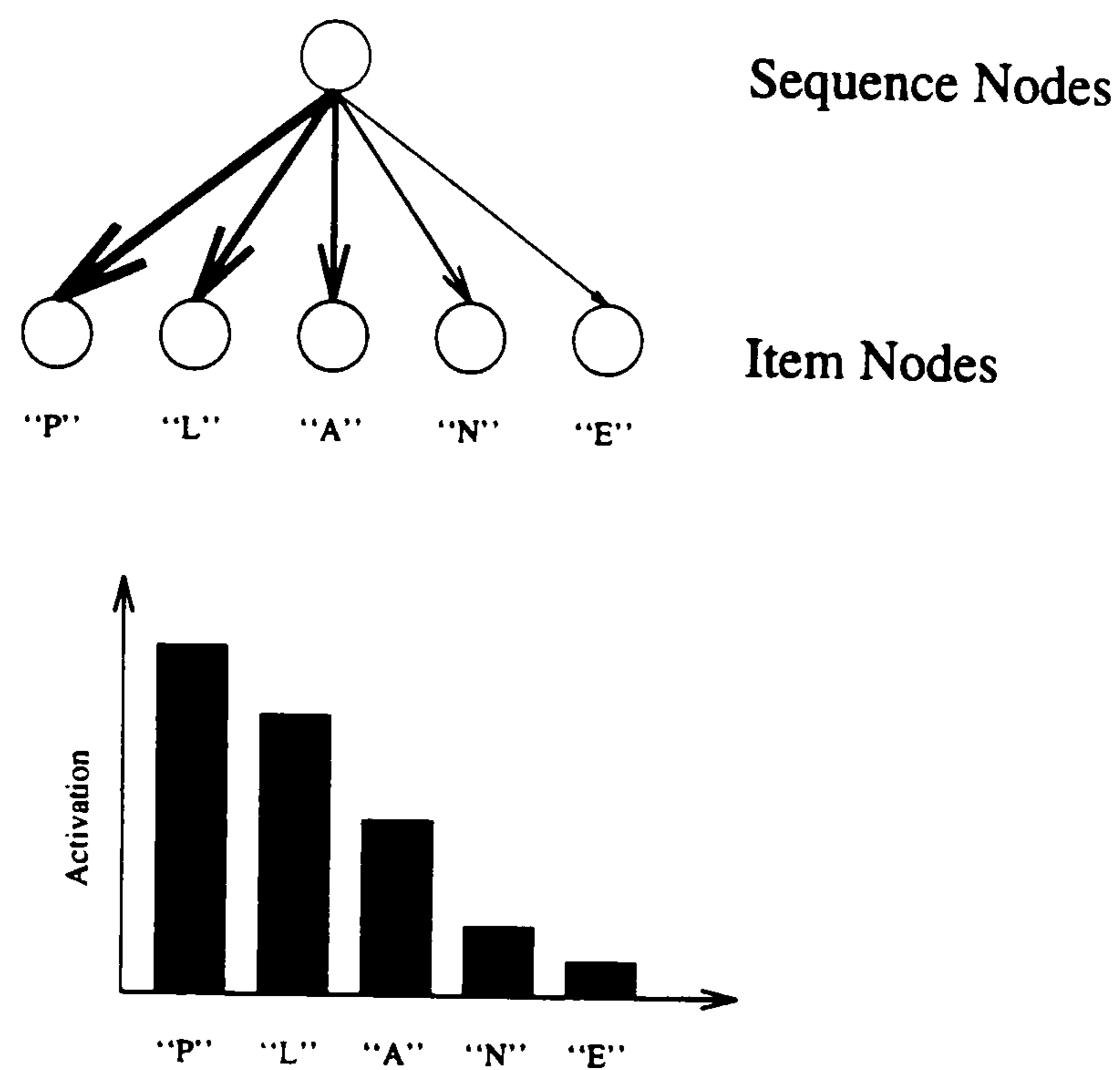


Figure 2.6: *Grossberg's (1978) model uses weights of varying strength between sequence and item nodes to set up an activation gradient, which is used to control serial behaviour.*

with empirical data. Their strong points derive from their use of parallel processing. The next section discusses models which build on the strengths, while avoiding the weaknesses.

## 4 P-I models

An alternative to I-I models is offered by models which suggest that serial order is represented not in terms of associations between items, but rather in terms of associations between items and serial positions.

The serial-ordering mechanism suggested by Estes (1972) and subsequently implemented in a connectionist network by Rumelhart and Norman combined the local representation of actions, with real activation values which were used to encode serial order. Much of the explanatory power of the approach derived from these innovations, and not from the use of a serial order representation based on item-item

associations. In fact, the sequence-specific lateral-inhibitory connections are only necessary if one assumes that the excitatory weights from the sequence nodes to the item nodes are equal. If this assumption is abandoned, the weights on the excitatory connections can be used to establish the activation gradient used to determine serial order. This was the approach suggested by Grossberg (1978) (see figure 2.6). Connections of varying strength from a sequence node encode serial order. When the sequence node is activated, item nodes with stronger connections will receive most input. This establishes an activation gradient across the item nodes. As in the Rumelhart and Norman (1982), the more active a node is the greater its influence on the output system. Concurrently active item nodes compete through non-specific lateral inhibition. The most active item node will exert most control over the output system, eventually causing the action it represents to be produced. The most active item node is then suppressed by feedback inhibition, allowing the subsequent actions to be produced in order.

As well as being somewhat simpler than the I-I accounts described above, Grossberg's approach to the representation of serial order overcomes several of the problems inherent in these schemes. In particular, since the lateral connections between item nodes are not sequence-specific, there is no interaction between different sequences involving the same items. An arbitrary number of sequences can be represented without any difficulty provided each is associated with a single node at the sequence level. This is shown in figure 2.7.

An additional advantage of Grossberg's approach is that it allows the rapid development of appropriate sequence-to-item node connections through exposure to a sequence using an unsupervised Hebbian learning rule. An unassigned sequence node must be continuously activated while a novel sequence is perceived. Perception of successive items in the sequence activates item nodes, which remain active as subsequent items are experienced. Weights between concurrently active sequence

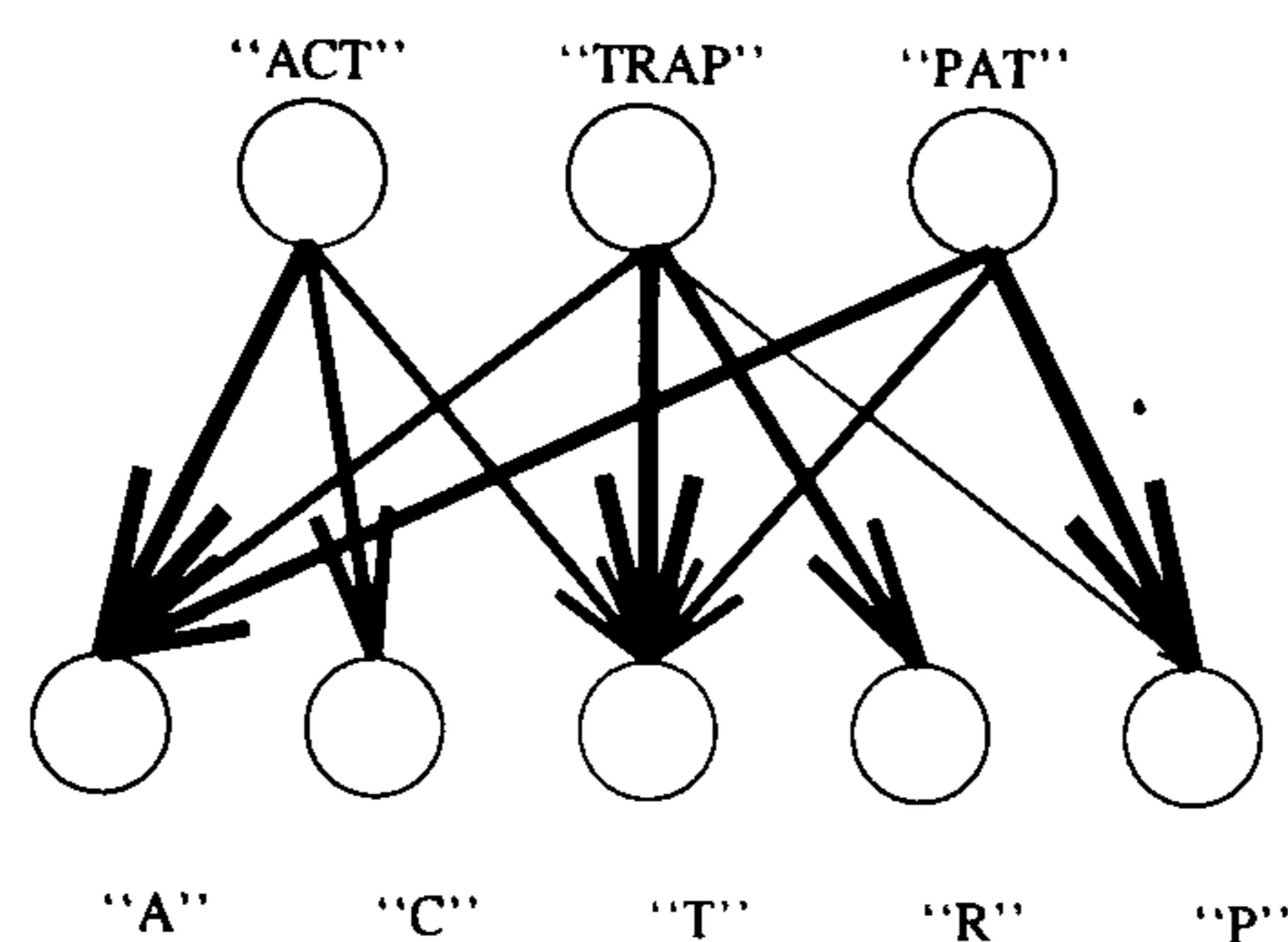


Figure 2.7: *Sequences coded using Grossberg's scheme (1978) do not interfere in the way that sequences coded as associative chains do. Here, for example, the weights used to represent the sequences "A,C,T", "T,R,A,P" and "P,A,T" do not conflict with one another.*

and item nodes are strengthened at each time step. Because items at the beginning of the sequence are active for the longest period, these nodes will develop the strongest connections from the continuously active sequence node. The strength of the connections thus represents the position of each item in the series.

However, one serious problem remains: the same connection strength cannot be used to represent more than one serial position. More recent P-I models have dealt with this problem by proposing that there is a dynamic, distributed pattern of activity at the sequence level. This approach has the advantage that sequences containing repeated items can be represented without using tokens; it is possible for an item to be linked with more than one serial position (different serial positions have different representations at the sequence level). The disadvantage is that it requires the specification of some additional mechanisms for the representation of serial position. Following Houghton (1990), such models are usually termed competitive queuing (CQ) models, since their implementation can vary a good deal, I will first present

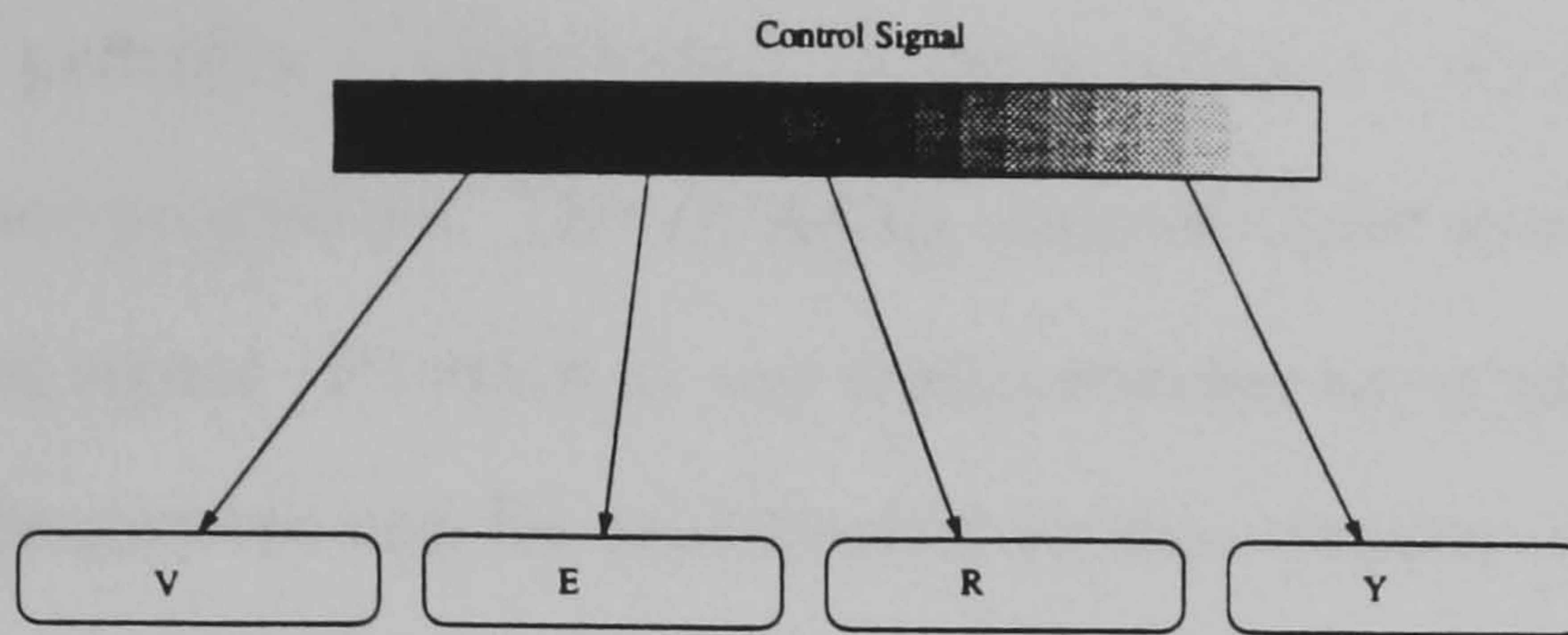


Figure 2.8: In competitive queuing (CQ) models, a time-varying control signal (or context) is used to encode serial position. Generally the signal, represented here by the shaded bar, changes gradually over time (from left to right), such that nearby states are positively correlated. Each item in a sequence is associated with a particular state of the control signal. Recovery of the control signal can be used to activate the items in sequence, but its temporally-autocorrelated nature results in competition between responses, especially if they are close together in the target sequence.

a general description of the mechanisms of competitive queuing, before describing some of the models in more detail.

#### 4.1 Competitive queuing

Houghton suggested that instead of representing each series of actions in terms of the links between a *single* sequence node and nodes representing those actions, a *group*

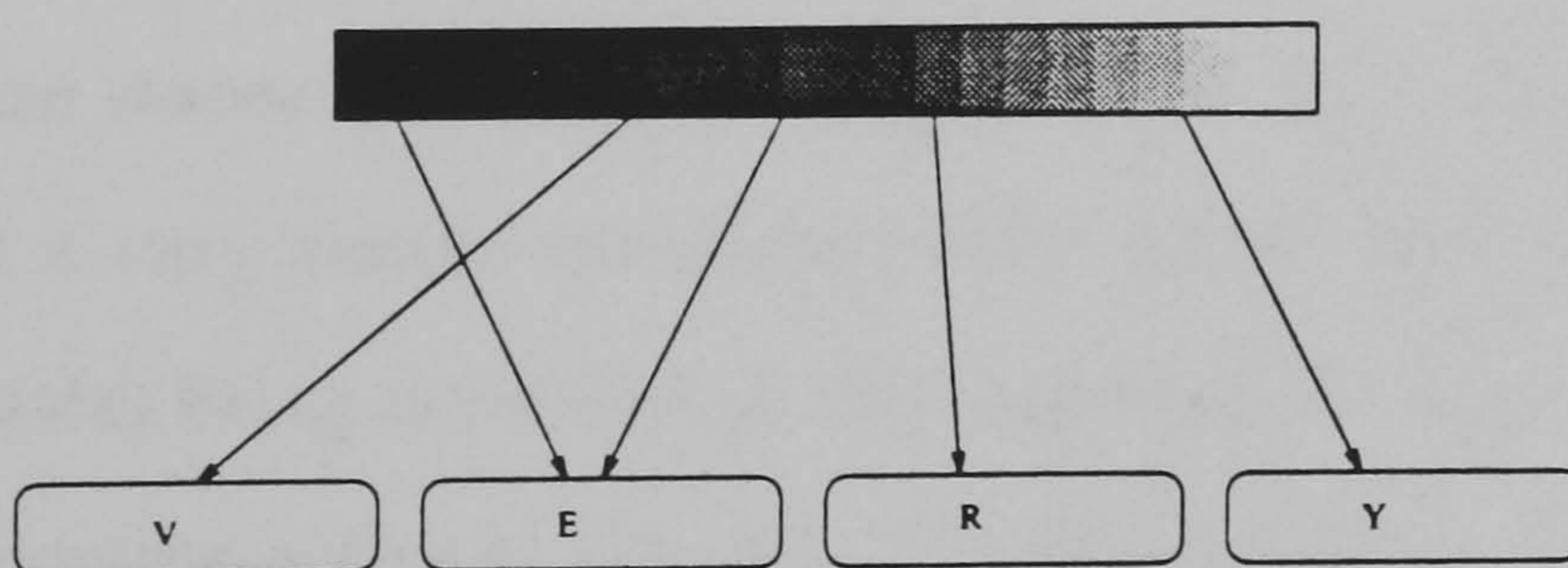


Figure 2.9: The representation of sequences containing repeats is not problematic for CQ models, since each instance of an item (e.g., each "E" in "E, V, E, R, Y") can be associated with a distinct state of the control signal.

of sequence nodes provides a distributed representation of serial position, which changes as a sequence progresses. The changing state of these nodes can be regarded as a dynamic control signal. Its state at any time provides an implicit representation of serial position. Sequences can be represented by associations between items and particular states of the control signal. This is illustrated schematically in figure 2.8. Because the control signal is dynamic, it is possible for an item to be associated with more than one position. In figure 2.9 "E" is associated with two distinct states of the control signal, one for each instance in the sequence "E, V, E, R, Y".

Like the Grossberg (1978) model, CQ models use Hebbian learning rules to develop an appropriate set of weighted connections as the net is exposed to sequential input. This involves activation (by some perceptual system) of item nodes corresponding to successive actions, one at a time. At the same time, the pattern of activation in the sequence nodes changes in a reproducible way. Weights between concurrently active sequence and item nodes are strengthened, forming an association between each item and a particular state of the control signal which was current when it was perceived. Whenever that state occurs, the item will become active. Recall can thus be accomplished by 'replaying' the control signal. The weights developed during learning propagate activation from the sequence nodes to the item nodes.

CQ models generally assume a form of control signal which is temporally-correlated. That is, the state of the sequence nodes changes gradually, so that nearby states are more similar than states which are well-separated in time. Houghton (1990), for example, suggested a very simple two-dimensional signal, with one sequence node (the initiator or I-node) being activated at the beginning of a sequence, and another (end or E-node) becoming active at the end. The self-similarity of the control signal associated with items which occur close to one another in a sequence means that they will tend to be active in parallel during recall. This has the effect of producing an activation gradient over the item nodes which is modulated over time as the

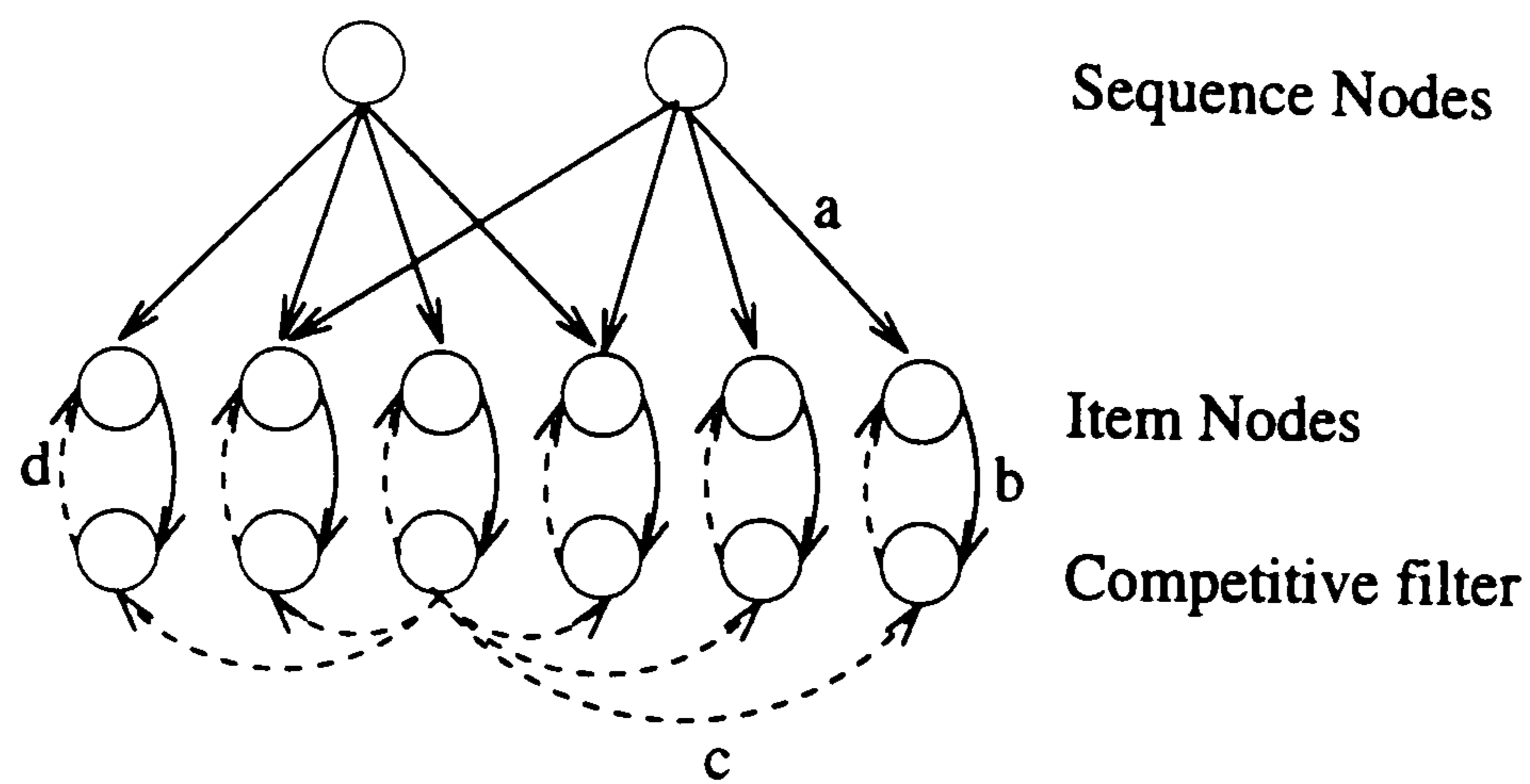


Figure 2.10: *Houghton's competitive queuing architecture (1990)*. Serial order information is represented in a distributed fashion in the learned weights (a) between sequence and item nodes. A temporally-modulated activation gradient is established during recall by 'replaying' the control signal experienced during learning. Item nodes feed their activation forward to the competitive filter via one-to-one excitatory connections (b). Lateral inhibition (c) between filter nodes suppresses all but the most active item node, and allows that item to be output. The selected item is then suppressed by means of one-to-one feedback connections from filter to item nodes (d).

control signal changes. The most active item at any time will be the one most closely associated with the current state of the control signal.

The activation gradient across the item nodes can be used to drive a competitive output process, like that visualised by (Estes, 1972) and (Grossberg, 1978) and others. Houghton, though, proposed that this competitive process was separate from the processes which activated the item nodes, taking place in a dedicated competitive filter, which provides a general mechanism for the resolution of response conflict. In the competitive filter (see figure 2.10) nodes representing the items to be sequenced are activated by one-to-one feedforward connections from the item nodes. Filter nodes suppress less active competitors through lateral inhibition until only one remains active. At this stage the item is output, and the corresponding item node is suppressed by means of inhibitory feedback connections.



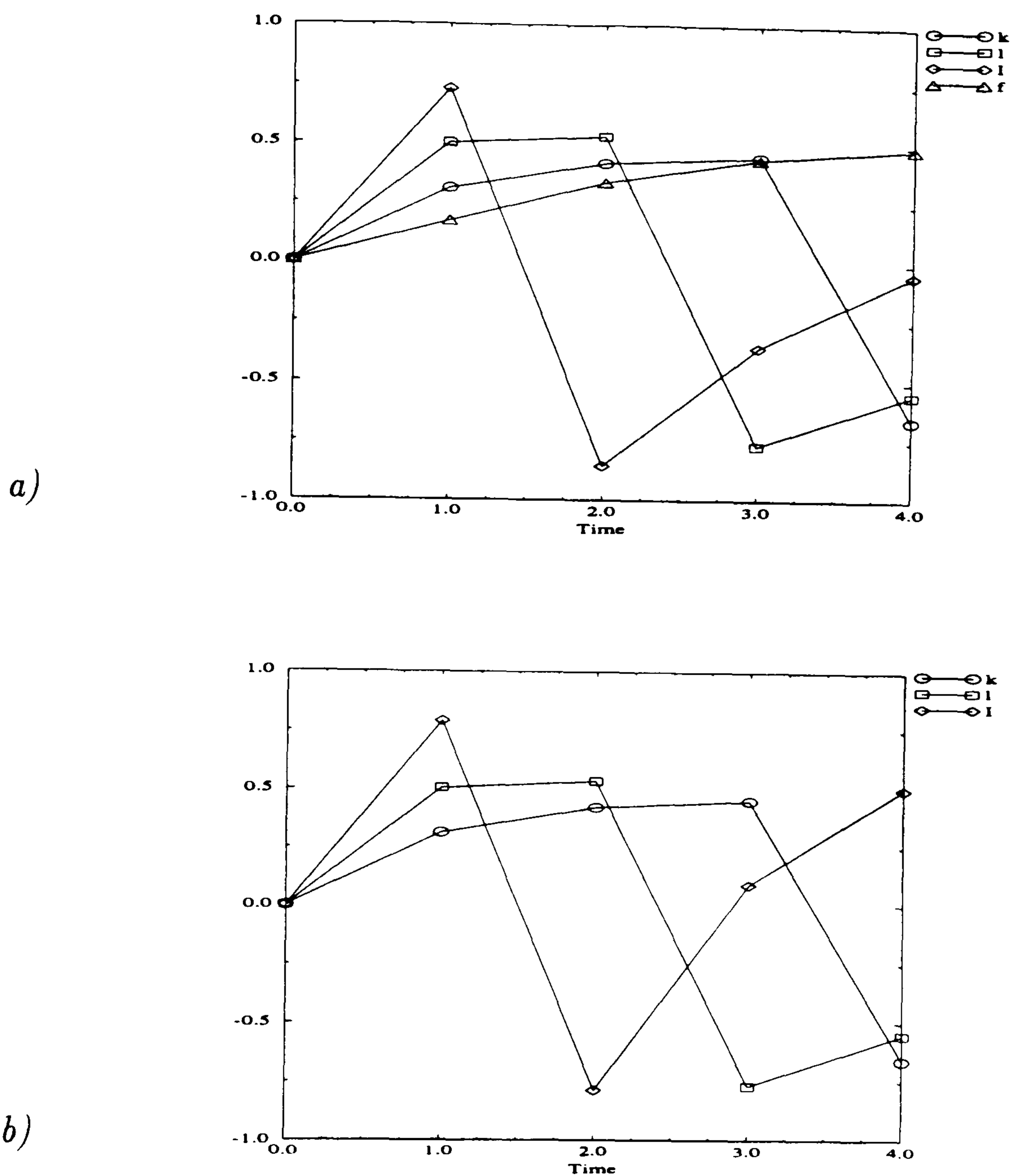


Figure 2.11: *The changing activation of a set of item nodes in a Competitive Queuing model similar to that described by Houghton (1990). a) Recall of a short non-repeating sequence ('cliff'). b) Recall of a sequence containing a repeated item ('click') showing the recovery from suppression of the item node representing /k/.*

The dynamics of the process are illustrated in figure 2.11a, Houghton (1990) which shows the activation of a set of item nodes as recall progresses. Because an item can be associated with more than one state of the control signal, its activation may be modulated by input from the sequence nodes throughout the recall process, so that a repeated item can recover from suppression. This is illustrated in figure 2.11b.

As in the Rumelhart and Norman (1982) model, ordering errors can occur when more than one action is strongly activated at the same time during recall. In this situation, any noise in the system can easily lead to the wrong competitor being selected – a transposition error. The more similar the control signal at two points in the sequence, the more competition between the items that occupy those positions, and the more likely it is that the items will transpose with one another. Typically, because of the temporally-correlated nature of the control signal, the substituting item will have occupied a position close to the one it replaces in the target list. The replaced item will remain active, and can thus win the competition for the next position, and this can often result in a complete exchange.

Competitive queuing has proved quite promising as a general theory of serial learning, spawning models of speech production (Houghton, 1990), spelling (Houghton et al., 1994), and serial recall (Burgess & Hitch, 1992; Glasspool, 1994; Burgess, 1995). The behaviour of CQ models, and the nature of the errors they predict is critically dependant on the nature of the control signal they use. Temporally-correlated control signals produce mainly local transposition errors, which are common in typing (e.g., Shaffer, 1976) and serial recall (e.g., Henson, Norris, Page, & Baddeley, in press). Additional factors constrain the nature of phonological ordering errors in speech production and these will be discussed in chapters 3.

Burgess and Hitch's (1992) model of the articulatory loop is of particular interest, because it represents the first attempt to apply a CQ architecture to serial

short-term memory. This, they hoped would extend the predictive and explanatory power of the articulatory loop model (Baddeley & Hitch, 1974; Baddeley, 1986). Hitherto, specification of the mechanism responsible for the maintenance of serial order information had been confined to a metaphorical reference to a tape loop of fixed length (Baddeley, 1986). Some of the most characteristic properties of verbal short-term memory could not be accounted for in terms of this metaphor. In particular it failed to explain the typical pattern of ordering errors that involve paired transpositions (of often adjacent items), the characteristic bowed shape of the serial recall curve, and the sigmoidal decline in lists correctly recalled as a function of list length. Burgess and Hitch (1992) set out to address this problem by developing a computational model that would provide an explicit serial-ordering mechanism: competitive queuing. They also wished to evaluate different approaches to serial order representation. They thus included a mechanism for item-item chaining, as well as a positional representation.

The model has a layered structure, with successive (feed-forward) layers representing input phonemes, words, a competitive filter, and output phonemes. During learning, activation of the nodes in the input phoneme layer activates word nodes (representing words containing the active phonemes), and these are associated with a dynamic control signal (represented by the context nodes) using Hebbian learning as in the Houghton model. Activation is fed forward, via the competitive filter to the output phoneme layer, where the constituent phonemes of the selected word are activated in parallel. The chaining mechanism involves strengthening feedback connections from phoneme nodes in one output pattern to phoneme nodes in the next input patterns.

At recall, the activation of a word node is dependent on both context and chaining input. The temporal cueing of responses using temporal context implements a competitive queuing mechanism similar to Houghton's (1990). The control-signal

used during learning is reproduced, producing a modulated activation gradient in the output word nodes. Response conflict is then resolved in a competitive filter. Inhibitory feedback from the competitive filter layer to the word layer ensures that words that have been output (subvocally rehearsed) are immediately suppressed. Burgess and Hitch made no attempt to model the mechanism responsible for the sequencing of phonemes within each word, which were instead represented as unordered patterns of activation in the phoneme layer; the winning word node simply activates its constituent (output) phonemes in parallel.

In contrast to Houghton's (1990) model, Burgess and Hitch (1992) used a control signal of arbitrary dimensionality. At each timestep (word presented) during learning, a random selection of context nodes (forming a small proportion of the total set) changed state (from being active to inactive or vice-versa). Because from one timestep to the next, the proportion of nodes changing state was small, the signal is temporally-autocorrelated. As in the Houghton model, the activation of items during recall depends on their temporal separation (during learning) from the current target. However, the high-dimensional control signal Burgess and Hitch used is potentially much richer in temporal information – the difference between temporally-separated states (as measured by, say, euclidean distance) can be much greater. As a result, the modulation of the activation gradient can be more specific, with less activation of items which are not immediately forthcoming in the sequence. Using the minimal two-dimensional control signal proposed by Houghton, a large proportion of list items are necessarily associated with both sequence nodes. The correlations between states of the control signal representing different serial positions are very strong for the middle part of long lists. Items from all these positions would compete with one another during recall – resulting in frequent errors for sequences longer than a few items. Houghton's model thus required a supervised 'practice' phase to learn such sequences. In the Burgess and Hitch model the number of items which could be retained in order is not constrained by the dimensionality of the con-

trol signal. Instead, Burgess and Hitch proposed that the capacity of short-term memory is limited by trace-decay; the strength of the learned weights decays exponentially over time. As the interval between the learning of a word and its recall increases (e.g., as list length increases), the input to the word nodes decreases. The effect is to reduce the advantage of each target item over its competitors, making the system increasingly vulnerable to noise as the list length increases. In line with the original articulatory loop model, rehearsal plays an important role in maintaining a trace in the face of decay. Rehearsal is modelled by going through the recall process while strengthening the learned connections to selected word nodes. Rehearsal thus directly counteracts the effects of weight decay.

In a number of simulations, the model showed many of the basic properties of human immediate memory. Crucially, the model's performance declined sigmoidally as list length increased and produced characteristic ordering errors such as exchanges. A phonemic similarity effect was also demonstrated. This property of the model results from the fact that during learning, a node in the word layer can be partially activated when an item phonologically similar to the word it stands for is presented. This means that during recall, a population of phonologically similar words will be associated with each state of the temporal context, and will compete to be output. The greater the degree of parallelism in the word layer, the greater the scope for the wrong item to be selected.

The Burgess and Hitch model had been expected to show the classic bowed serial recall curve. The reason for this expectation was as follows. In a CQ model, each item is associated with a particular state of the control signal. The states corresponding to different positions correlate to some degree with their temporal neighbours, and this leads to competition between neighbouring items. At either end of the list, though, there are fewer neighbouring items which can be activated in parallel. Thus these list positions gain an advantage through their distinctiveness.

However, the original implementation of the model did not show any recency effect. The authors believed that this was because random changes in the control signal often lead to spurious autocorrelations between temporally-separated states of the control signal<sup>1</sup>. They proposed a simple modification to the control signal to prevent this undesirable effect. Instead of randomly changing at each timestep, the active elements of the control signal would change systematically, so that nodes associated with the beginning of the list would not be active at the end of the list. The adapted control signal retained the temporal-correlation between nearby states necessary to explain local transposition errors. Recently, Burgess (1995) showed that the modified model produces a bowed recall curve which closely matches that found in experimental studies.

The use of the connectionist paradigm allowed the authors to experiment with different blends of chaining- and position-item associations operating in parallel. Through this process they came to the conclusion that explicit chaining associations produced behaviour which was uncharacteristic of human performance, a view supported by recent empirical work (Henson et al., in press) and the theoretical considerations developed earlier in this chapter and elsewhere (Lashley, 1951; Houghton & Hartley, in press). The use of chaining has been abandoned in more recent work based on the Burgess and Hitch model (Burgess, 1995).

Although the Burgess and Hitch (1992) account was successful in explaining some of the more important characteristics of verbal short-term memory, the representations of phonological forms used were long-term lexical ones that had to be learned before short-term learning and recall could take place. As I have pointed out (see chapter 1) the ability to recall unfamiliar phonological forms (novel or nonsense words) is

---

<sup>1</sup>This would not have been the case had the number of context nodes been large enough. With a very large number of context nodes, it is very unlikely that a particular node will be first deactivated and then reactivated later on in a sequence.

---

quite basic to the short-term memory system, and appears to play a role in long-term learning. It is therefore essential that any model of short-term memory be capable of handling such stimuli.

## 5 Summary and Conclusions

Competitive queuing models provide a particularly promising approach to the problem of serial order. Unlike chaining models, they provide a good account of ordering errors in tasks such as typing and serial recall, and can readily handle sequences containing repeated items. Furthermore they do not demand powerful supervised learning algorithms, but can learn sequences in a single exposure using unsupervised learning algorithms. This capacity is essential for any viable model of short-term memory. The applicability of CQ to short-term memory has been demonstrated in recent modelling work by Burgess and Hitch. For these reasons, the CQ approach will be adopted in the model developed here.

Although CQ provides a useful framework for modelling of serial order, previous models have not considered how serial ordering might interact with task specific constraints. The next chapter explored empirical work showing how linguistic constraints on syllable structure affect phonological errors in the immediate recall of nonwords and lists of nonwords. A major challenge for the current work will be to integrate these constraints into an architecture consistent with Burgess and Hitch's CQ model of the articulatory loop.

### **3. Empirical background: serial order and memory for speech**

---

This chapter reviews some experimental studies which provide the empirical constraints on the model of short term memory described later. The first section describes memory experiments which show how stimulus familiarity quantitatively affects performance on short-term memory tasks involving verbal materials. In the second, qualitative analyses of phonological errors made in the recall of nonwords are discussed. The third section introduces data from psycholinguistic studies of speech errors, and highlights the similarities between these and errors of nonword recall. Finally the important points in the data are summarized.

#### **1 Stimulus familiarity and recall**

The familiarity of the items used in a test of serial recall affect subjects' performance on the task. The quantitative difference between memory for familiar and unfamiliar



## 1. STIMULUS FAMILIARITY AND RECALL

---

stimuli is most clearly established by a direct comparison of memory span for words and nonwords. Hulme, Maughan and Brown (1991) conducted a study in which subjects were asked to serially recall lists of composed of either words or nonwords. Span was determined by presenting subjects with lists of increasing length, and recording the maximum length for which serial recall was error-free for four lists of the same length. After an error occurred, longer lists were presented, with subjects scoring 0.25 for each list correctly recalled at each length, until no lists were recalled correctly. Span was found to be shorter for nonwords than for words of the same spoken duration.<sup>1</sup> For both words and nonwords, span was found to vary linearly with the speech rate applicable to the items concerned. The slopes for the two linear functions did not differ significantly, while the intercepts did. The implication is that the value of the constant  $c$  in equation 1 is dependent upon stimulus familiarity. The difference due to familiarity is clearest for monosyllabic items where articulation rate is approximately equal for words and nonwords; on lists of familiar monosyllables span was about 5.0, whereas for the novel stimuli it was around 3.5.

Hulme et al. suggested that the recall of familiar items was supported by long-term phonological representations which were not available in the recall of the novel items, and that this additional information was responsible for the difference in the intercepts; both word and nonword recall was supported by a mechanism responsible for the shared gradient of the two functions. They followed Baddeley (1986) and others in postulating that this mechanism employs a non-lexical code and an articulatory control process.

Hulme et al. (1991) did not describe their subjects' errors. However, a comparison of findings from other experiments indicates that the principle effect of a reduction

---

<sup>1</sup>Non-words were found to be articulated more slowly than words, so that the articulation rate for five-syllable words (e.g., hippopotamus) was comparable to that for three-syllable non words (e.g., bepavit).

## 1. STIMULUS FAMILIARITY AND RECALL

---

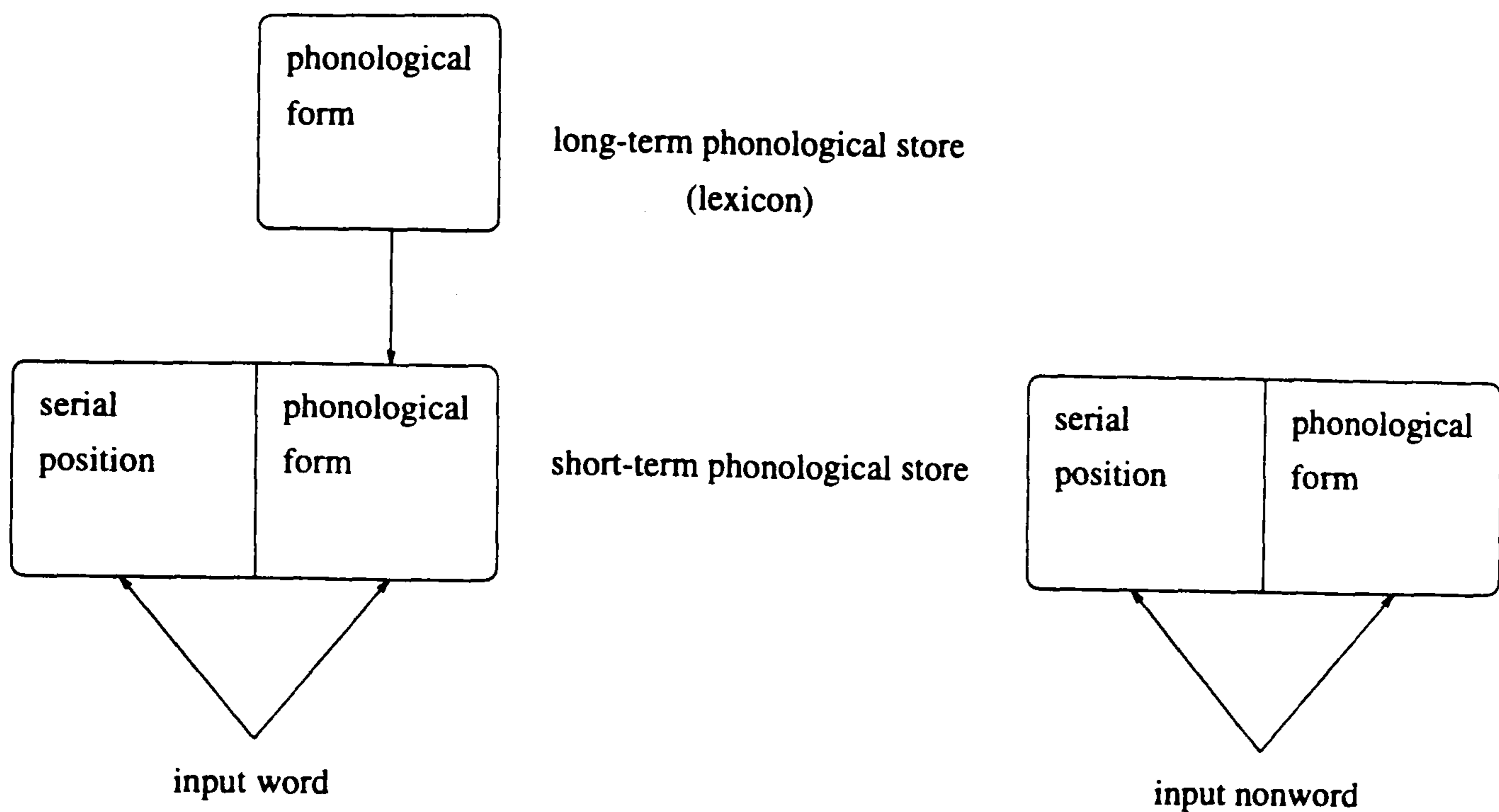


Figure 3.1: *The robustness of the short-term representation of the form of a familiar item is enhanced by the availability of a long-term representation. This long-term knowledge cannot be exploited in the representation of novel forms.*

in stimulus familiarity is to increase the rate at which phonological errors occur. When subjects are required to serially recall a list of *familiar* words, the majority of observed errors involve the misordering of intact items. Sometimes list items are omitted at recall. A subject may occasionally generate a word which did not appear in the target list – such items will typically be drawn from the experimental ‘vocabulary’ where it is a clearly limited one (e.g., digits or letters of the alphabet). Phonological errors, however, are rarely reported. It seems that where the listed items are familiar words, serial recall is virtually free of such errors, in normal adults at least. This is not the case when the target sequence is a list of *unfamiliar* syllables. Under these conditions, normal subjects make a large number of phonological errors, the majority of which involve the recombination of *phonological fragments* of the target items. Simple misorderings of the target items are relatively infrequent (Ellis, 1980; Treiman & Danis, 1988). These results are discussed in more detail in the next section (‘Phonological constraints on errors’).

Phonological errors occur in addition to item-ordering errors which are found in

## 1. STIMULUS FAMILIARITY AND RECALL

---

both word and nonword recall. This observation suggests that phonological form and item-order information are separate components of the representation of a word (or nonword) in STM. This leads to the following interpretation of the findings (see figure 1): phonological errors arise as a result of the failure of the component of the short-term store responsible for retaining the detailed forms of the target items. This sub-system is redundant in the word recall task, because the forms of the targets are already well-learned; the trace of an item's phonological form in the short-term store is augmented by 'top-down' input from the lexicon. This additional contribution of long-term knowledge apparently provides a representation of the word's form sufficiently robust to prevent phonological errors from occurring at all frequently in normal adults. This is not surprising, since it is presumably the same long-term knowledge that also supports the production of familiar forms in spontaneous speech. In the serial recall of word lists the essential problem for the short-term memory system is to hold information about the serial order of the items, which cannot, generally speaking, be supplemented by long-term knowledge<sup>2</sup>. The capacity of the short-term store to retain such information is limited, and as the material to be stored increases in duration, item-ordering errors are observed in recall. In nonword recall, the item-order information must also be retained, but there is an additional problem: unsupported by the lexicon, recall relies on a short-term representation of the phonological form of the input.

This view of the role of long-term phonology in short-term memory is supported by another recent experiment, showing how phonological memory is adversely affected by a deterioration of vocabulary brought on by disease. Patterson, Graham, and Hodges (1994) describe the phonological errors of patients suffering from semantic

---

<sup>2</sup>If familiar sequences e.g., "F, B, I" or "7, 4, 7" appear in stimulus lists, long-term knowledge *can* be used to enhance recall. In some practiced mnemonists, such 'chunking' strategies can be exploited to enable the single trial learning of extremely long sequences (see e.g., Ericsson, Chase, & Falloon, 1980).

---

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 gradually access to familiar words is 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. Patterson et al. concluded that use of the (long-term) phonological lexicon depends on the availability of semantic knowledge which has been disrupted in these patients. When these *long-term* representations are no longer available, the phonological errors characteristic of memory for nonwords are observed.

A number of studies have been carried out on subjects with acquired and developmental disorders, who have particular problems with nonwords (compared to familiar words) in a range of tasks assumed to involve working memory. In these cases, the nonword-specific deficit is generally ascribed to the failure of a subsystem dedicated to the retrieval and/or output of unfamiliar phonological forms. However, it is noticeable that, in almost every case, the deficit is associated with sub-normal digit span.

Caramazza et al. (1986) report a patient, IGR (digit span 3), with a selective deficit in nonword reading, writing and repetition. Bisiacchi et al. (1989) report that a patient RR (digit span 3) with no apparent difficulty with phonological processing read, wrote and repeated fewer nonwords correctly than an age and IQ matched control group. Bisiacchi et al. (1989) conclude that RR's selective deficit for nonword processing is due to damage to a dedicated phonological buffer, as postulated by Caramazza et al. (1986).

As well as acquired deficits, there are examples of developmental syndromes char-

acterized by particular difficulties with tasks involving the retention and processing of nonsense words. Language-disordered children, whose non-verbal intelligence is within normal limits, show poor performance on tests of vocabulary reading and language. They are also distinguished from children matched for age and non-verbal IQ, and from younger children matched for language abilities, by markedly poorer performance on tests of nonword repetition (Gathercole & Baddeley, 1990).

Campbell and Butterworth (1985) report the case of a highly literate subject, RE (digit span c. 3.5), who had grave difficulties with the reading and writing of nonwords, whilst exhibiting unimpaired performance on analogous tasks involving words. This subject was described as phonologically dyslexic and dysgraphic. The developmental phonological dyslexia of another subject, JM (digit span 5), has been shown to be associated with impaired nonword repetition (Hulme & Snowling, 1992). In this case, the authors noted that JM had previously been able to discriminate pairs of phonologically similar nonwords. They argued that the subject's difficulty in repetition was best characterized as an output rather than a storage deficit since the discrimination task had storage demands that are similar to or greater than the repetition task. However, this assertion is questionable. Whilst it may be that a greater temporal span of storage is required for the discrimination task, the quality of the representation required for the two tasks also differs. A quite degraded representation of the input may nonetheless contain the information necessary to make a same-different judgement, and if it does not the chances of an error are still only 50%. Repetition clearly demands that the entire phonological form can be retrieved accurately. The distinction is analogous to that between recognition and recall: the subject need retain only a few relevant cues to recognize a stimulus when it is re-presented, but a much fuller representation is required for recall.

These neuropsychological data have been used to motivate a number of theoretical proposals regarding the differences between word and nonword processing. Re-

searchers in general have been unwilling to attribute disorders like those described above to a general failure of short-term phonological memory, preferring instead to postulate the existence of specialized buffers and/or output processes for unfamiliar materials. For instance, Caramazza et al. (1986) interpreted their findings in terms of the operation of a phonological (output) buffer, which, they suggested, is used to hold phonological representations temporarily while readout procedures are carried out in nonword reading, writing and repetition.

The existence of this specialized buffer is demanded, they argue, because readout processes require access to a span of phonological information in order to create an articulatory (or graphemic) representation of the unfamiliar form demanding a special form of storage. Caramazza et al. (1986) propose that, in contrast, there are direct routes which support *lexical-articulatory* and *lexical-orthographic* translation without buffering. However this view is somewhat contradicted by their suggestion that the phonological (output) buffer *is* implicated in errors of sentence production, since such errors typically take the form of *phonemic* substitutions (see section 3) – a category of error which they regard a characteristic failure of the buffer. IGR made a large number of single-phoneme substitutions in the reading, writing and repetition of nonwords and Caramazza et al. (1986) suggest that these occurred because the phonological representation of the target item had been underspecified (perhaps through rapid decay, resulting from of damage to a phonological buffer), but that enough *structural* information may be retained to generate an error bearing a phonological similarity to the target. Essentially the difficulty with Caramazza et al's account is that the interaction between articulatory and storage systems is not well-defined. It is thus hard to see how each contributes to the observed patterns of error.

Bisiacchi et al. (1989) also conclude from their case study that RR's selective deficit for non-word processing was due to damage to a phonological buffer of the type described by Caramazza et al. (1986). They do, however, recognize that it is quite

usual for disorders of non-word processing (including phonological dyslexia and dysgraphia) to be associated with short term memory impairment (as assessed by standard span tasks), and speculate as to whether or not the two systems overlap. They are forced to conclude that they do not, since at least one patient, PV, described as having a specific impairment (Basso et al., 1982), could read nonwords correctly, despite being unable to write or repeat them (Vallar, in a personal communication to Bisiacchi et al.). It is important to recognize, however, that the demands of reading on phonological memory are less than those of repetition or writing, since the subject has, at the time of making the response, access to an orthographic representation of the nonword. In writing to dictation or repetition, by contrast, it is necessary for the subject to form a temporary internal representation of the phonological form, unsupported by any enduring external representation.

The aversion to a more global account of disorders of nonword processing seems to stem from an assumption that familiar and unfamiliar words have short-term representations which are equally vulnerable to disruption in normal subjects. Thus it is argued that if performance on one type of stimulus is selectively affected by a disorder, a dissociation between separate word and nonword subsystems has been demonstrated. However, the evidence from Hulme et al.'s (1991) study suggests that this assumption is not valid: if the capacity of short-term memory is reduced, the representation of familiar words will still be supported by long-term phonological knowledge. There exists, therefore, the potential for global damage (or developmental abnormality) in the short-term phonological store to differentially affect storage of familiar and unfamiliar materials, while the retrieval processes common to both types of stimulus remain intact. For familiar materials, the short-term storage of phonological form is unlikely to be affected, since it is supported by *enduring* (and presumably undamaged) lexical knowledge. The retention of serial order information cannot be supported in this way. This being the case, we would anticipate a decline in the capacity to store item-order information in patients affected by such

disorders. The co-occurrence of subnormal span and nonword repetition deficit is certainly marked in the cases reviewed above, yet it remains unexplained by the model advanced by Caramazza et al. (1986), which would presumably require that the span deficit arises independently of the nonword repetition deficit. Their dual-locus account of the syndrome is less parsimonious than the single-locus account outlined above.

## 2 Phonological constraints on recall

As indicated above, subjects recalling sequences of nonwords are prone to make phonological errors which are not at all common in the recall of familiar words. The erroneous responses almost always share phonemes with one or more of the targets. Often, they simply recombine fragments of the listed items. Further investigation has shown that 'recombination' appears to be subject to linguistic constraints. Treiman and Danis (1988), for example, provide a detailed qualitative analysis of error types in nonword serial recall. Subjects were presented auditorily with lists of 6 CVC nonsense syllables, which they then recalled in serial order. Averaged over serial positions, 42% of responses were correct. Subjects were encouraged to guess rather than make "don't know" responses, which accounted for 3% of all responses. Of the remaining errors (54% of responses), 96% had the same consonant-vowel structure (CVC) as the target syllables. Relatively few of these were item order errors (about 6% of all responses). Many (28% of responses) were made recombinations of phonemes from different list items. In addition, many of the CVC errors (17% of responses) included non-contextual substitutions (phonemes which did not appear in any of the list items).

A similar pattern of results was observed for syllables with different shapes (CCV and VCC). In each case, though errors were numerous, the great majority preserved



## 2. PHONOLOGICAL CONSTRAINTS ON RECALL

---

the CV-structure of the target syllables. The fact that responses tend to maintain the CV-structure of the target items, despite the fact that many substitution errors occur, indicates that consonants and vowels are constrained not to substitute for one another. Substitutions clearly greatly outnumbered any insertion or deletion errors which occurred; many of the responses combined a single phoneme from one target with two from another. Similarly, in nonword repetition single phoneme substitutions are the most frequent category of error (Caramazza et al., 1986; Gathercole et al., in press). A number of studies have shown that where substitutions do take place, consonants are more vulnerable than are vowels. For instance, Ellis (1980) found that vowels were recalled in the wrong serial position in 6% of responses, consonants in 23%.

Phonemic substitutions are further constrained. The more features two phonemes share, the greater the likelihood that they will substitute for one another (Ellis, 1980). Ellis refers to this as a feature similarity effect, although some nonword repetition data suggests that some features are more critical than others in determining the likelihood of one phoneme replacing another. Caramazza et al. (1986) present data showing that in 82% of IGR's substitutions the substituting phoneme shared a particular feature (manner of articulation) with the target phoneme it replaced. Bisiacchi et al. (1989) report similar findings from their analysis of RR's repetition errors.

In contextual substitutions the target and error segments are likely to share the same syllable position in the target lists, so that, for example, syllable-initial segments tend to replace other syllable-initial segments (Ellis, 1980; Treiman & Danis, 1988). The error segment is likely to have its origin in an item close to the point at which it appears in the error (Treiman & Danis, 1988). Vowels are particularly likely to maintain their serial position in errors.

## 3 Constraints on speech errors

Unintentional slips in spontaneous speech have long been identified by psychologists as providing useful data for the elaboration of the mechanisms supporting speech production. Freud (1901), for example, viewed speech errors as providing valuable insights into the character of the unconscious. More recently, however attention has focussed on the elaboration of the lower levels of the production system, in particular on the processes intervening between a semantic specification of a target utterance, and its phonetic specification and realization (see e.g., Shattuck-Hufnagel, 1979; Butterworth, 1992). The term phonological encoding is often applied to such processes, but it will be avoided in this thesis as it is confusing in the context of models of learning and memory. Instead, I will refer to such processes as phonological retrieval processes. Many speech errors can be understood as occurring at the level of phonological retrieval. At issue here is the extent to which the mechanisms underlying such errors are common to the retrieval of phonological forms from short-term memory.

The last section established that certain constraints apply to phonological errors in nonword recall. Errors typically involve the *substitution* of one or more *phonemic units*, for *phonologically-similar* ones, often originating *nearby in the target utterance* (list to be recalled). As well as being related by articulatory features (in particular manner of articulation), where an error has a contextual origin, source and error segments tend to have occupied the *same position in different target syllables*. In this section, the aim is to compare these constraints with those that have been found to operate in slips of the tongue. In the short-term memory experiments described above, the nonword stimuli are purely phonological without syntactic role or semantic association. With this in mind the discussion will concentrate on slips involving the phonological rather than, for example, syntactic or morphological units. Such

### 3. CONSTRAINTS ON SPEECH ERRORS

---

errors occur rather infrequently in the spontaneous speech of mature, normal speakers. In recordings of normal conversation, Garnham, Shillcock, Brown, Mill, and Cutler (1982), for example, identified 62 such errors per in around 200,000 words produced in conversation by several speakers being recorded without their knowledge (Svartvik & Quirk, 1980); Shallice and Butterworth (1977) report a rate of 1.6 errors per thousand words. Despite this some fairly large collections have been made and analyzed. However, the scarcity of data, and problems with sampling methodology (see below) have lead some researchers to develop experimental paradigms for eliciting speech errors. Others have concentrated on slips made by speakers who have particular clinical problems which make them more error-prone.

#### 3.1 Errors in spontaneous speech

Typically analyses are based on collections of errors built up by interested researchers. This methodology can and has been criticized (Cutler, 1982) on the grounds that because only the most salient errors are likely to come to the researchers' attention, there is a danger that the sample of recorded errors will be unrepresentative. Another general problem with speech error data is that the reported speech will have been filtered through the perceptual system, which is known to exert a correcting influence on anomalous utterances. Nonetheless data collected in this manner do show some striking regularities, which it would be difficult to account for purely in terms of sampling biases. Even if we are quite conservative in our treatment of the data, some useful conclusions can be drawn.

It is a property of speech errors that they can often be characterized in terms of the interaction between two units, one, the target, being the intended segment, the other, the source being a nearby segment from the intended utterance which interacts with the target segment. Thus in an error like "reading list → leading list", the target

### 3. CONSTRAINTS ON SPEECH ERRORS

---

segment /r/ has been replaced by /l/, the source of which is regarded as being the onset of "list". Constraints on ordering errors in speech describe the relationship of target and error segments – their linguistic characteristics and their positions in the intended utterance. These contextual errors have a similar character to the most frequent type of short-term memory errors, in which fragments of target nonwords are recombined. In both spontaneous speech, and the serial recall/repetition of nonwords, non-contextual errors also occur, where no source for the error segment can be identified within the target utterance. These, too, appear to be governed by linguistic principles.

The interacting units in errors of spontaneous speech generally coincide with the units of linguistic grammars: phonemes, morphemes, words etc. This has been widely taken as implying that linguistic units have a psychological role in the representation and production of speech. Generally, interacting units are of the same size and level of linguistic description. Thus, phonemes interact with phonemes, words with words and so on. This shows that ordering mechanisms are vulnerable to errors within but not between levels of linguistic analysis. This is suggestive of the involvement of a hierarchical sequencing mechanism, in which units at different levels of a linguistic hierarchy (for example syllables and phonemes) are ordered separately (Shattuck-Hufnagel, 1979).

Even given the reservations expressed above regarding sampling methodology, collections of sound errors in spontaneous speech leave little doubt that, as in non-word recall, the great majority involve the interaction of phonemic units. Shattuck-Hufnagel (1979) found that 70% of her collection of English speech errors were single phoneme errors (substitutions, deletions, additions etc of a single phoneme.) Nootboom (1969) in a corpus of Dutch speech errors found that 89% involved phonemic units. A sizable proportion (about 10-20% of errors) involve units longer than one phoneme, but shorter than a syllable. Errors involving entire syllables are rare. In

### 3. CONSTRAINTS ON SPEECH ERRORS

---

some rare cases, an error can be described in terms of the movement of featural units in the intended utterance, e.g., "Clear blue → glear plue". However, since such errors account for a very small percentage of recorded errors, and can always be understood in terms of one or more phonemic errors (i.e., in "glear plue" the /g/ and /p/ are independent non-contextual phoneme substitutions), it is debatable whether the interpretation of such errors as involving the movement of featural units is appropriate.

Broadly speaking, interacting units tend to share linguistic characteristics (Shattuck-Hufnagel, 1979). So far as phonological errors are concerned this means that they are phonologically similar. For example the /r/ and /l/ segments in the above substitution are both single phonemes, they are also both liquids, and they share a number of articulatory features. The resulting utterance does not violate the phonotactic rules of English, in this instance, 'leading' is a well-formed English word. In fact there are very few recorded instances of errors which have violated the phonotactic rules of the speaker's language. More specifically, consonants "almost never" (Shattuck-Hufnagel, 1979) interact with vowels and vice-versa. In an analysis of complete spoonerisms, consonants never exchanged with vowels (Motley, 1973).

Beyond effects of consonant-vowel category, the precise role of phonological similarity in constraining the interaction of segments involved in errors is again a little more difficult to establish. Boomer and Laver (1968) for example found no evidence for a feature similarity effect in their analysis of a small corpus of sound errors. Nootboom (1969), Mackay (1972), Goldstein (1977, cited in Shattuck-Huffnagel, 1979) and Shattuck (1975, also cited in Shattuck-Huffnagel, 1979) on the other hand have all reportedly found feature similarity effects. As in the errors of nonword repetition, the phonological principle governing errors, seems to be influenced more strongly by similarity along some articulatory dimensions than others. For instance, in the Mackay (1972) study, using data collected by (Meringer & Mayer, 1895), the

### 3. CONSTRAINTS ON SPEECH ERRORS

---

effect of similarity was found to be confined to features defining manner of articulation (nasality and voicing), but not those concerned with place of articulation. The bulk of the evidence clearly suggests that some form of phonological similarity effect operates in phonemic errors, but it may be that the featural specification used by linguists is not an ideal description of phonological similarity for this domain.

As in errors of nonword recall, interacting units tend to occupy comparable positions within the target utterance, in spontaneous speech errors. Syllable initial phonemes, for example, tend strongly to substitute with other initial phonemes as in the above example. Shattuck-Hufnagel (1979) found just 4 examples from a collection of 211 between word exchanges, where the syllable position of the exchanging segments was not comparable. (Baars, 1976) found that 90% of exchange errors involved units from the same position (either initial, medial or final) in different syllables. Another implication of this constraint is that sound errors do not often involve the interaction of units from within the same syllable.

In many errors, as in the "leading list" example, there is an anticipatory component: one segment appears in advance of its intended position in the utterance. (Lashley, 1951) made the point that such anticipatory errors suggest that ordering processes act on a substantial span of the intended utterance (elements of which are primed or available for output before their articulation), and not just a single phoneme or word at a time. At the same time, though, it is equally clear that the span of planned speech on which ordering processes act is finite, perhaps limited to a few words or syllables. In spoonerisms for example, the exchanging segments are from words close to one another, and often adjacent in the intended utterance.

#### 3.2 Experimentally elicited slips

Slips of the tongue can be elicited experimentally. The procedure involves the visual presentation of pairs of words. A target pair e.g., “bad goof” is preceded by several bias pairs which are silently read by the subject. The bias pairs are presented essentially to prime production of the incorrect sequence of initial consonants. This can be achieved by transposing the initial consonants from the targets in the bias pairs e.g., “get booze”, “gave book” (Baars & Motley, 1974; Motley & Baars, 1975). When the subject is required to articulate the target pair, the initial consonants are typically often exchanged in a complete spoonerism: “gad boof”. These experiments allow normal subjects’ speech errors to be studied under controlled laboratory conditions. But perhaps just as importantly, they also represent an important conceptual bridge between data supposedly concerned solely with output processes in speech production (e.g., Shattuck-Hufnagel, 1979), and data concerned exclusively with the elaboration of storage mechanisms (e.g., Treiman & Danis, 1988). Although the paradigm is normally regarded as investigating the mechanisms of speech production, it also involves short-term storage, since subjects are required to respond to the target pair after a short delay during which the stimuli are not present.

A further experiment (Baars & Motley, 1976, experiment 1) indicated that complete spoonerisms could be induced simply by priming the second initial consonant of a pair. For example, an error like “gad boof” can be elicited by priming subjects with bias pairs like “get rake”. In fact the stimuli used Baars and Motley (1976) were *nonwords*. The authors note though that an identical procedure can be shown to produce phonemic exchanges with lexical stimuli. However, it is clear that in this experiment the representation supporting output was a short-term one. The authors do not report any qualitative difference in the results they obtain for words and nonwords, and presumably concur with the argument put forward here that the output processes supporting the recall of familiar and unfamiliar forms overlap with

#### 4. SUMMARY

---

speech production processes at the level of phonology. They do, however, note that the effect of using unfamiliar stimuli is to *increase* the rate at which spoonerisms occur in line with the interpretation of the short-term memory data described in section 1.

A further experiment (Baars & Motley, 1976, experiment 2) showed that by introducing a degree of confusion as to the order in which the items in stimulus pair should be articulated, it was possible to induce spoonerisms in the absence of any phonological priming. Together the results of the two experiments were interpreted as suggesting a phonological retrieval process in which serial order was specified at at least two levels, the phoneme level, and the word (or syllable) level. Errors were attributed to conflicts between the two levels of sequencing, brought on in this case by experimental manipulation, but arising naturally in the production of speech. For instance, they suggest that in adjective-noun pairings such as "bad goof" the adjective acts to modify the meaning of the phrase, rather than fundamentally changing its meaning. Thus it may be expected (they argue) to be specified later in the production process. Meanwhile the syntactic rules of English suggest that it should precede the noun. Whatever the merits of this argument, the suggestion that ambiguity about serial order at one level can affect serial order at a lower level will be central to the model of short-term memory developed in chapter 5.

## 4 Summary

There is a good deal of qualitative similarity between phonological errors in spontaneous speech, and in short-term memory tasks involving nonwords. In a computational sense, both tasks have a great deal in common. They both clearly entail the interaction of retrieval processes with stored phonological forms to translate the internal representation of a target utterance into a sequence of articulatory actions.



#### 4. SUMMARY

---

There are two essential differences between the two tasks: firstly, the generation of spontaneous speech clearly involves pragmatic, semantic, and syntactic processing which is not required by the experimental tasks. Secondly, and more importantly in spontaneous speech, along with serial recall tasks employing familiar materials, long-term representations of the target forms are available. In such circumstances, phonological errors are rare. Where they have been analyzed, however, they have been found to be of a similar nature to those described in the recall of nonwords.

There seems little reason to suppose that the retrieval processes involved in the recovery of familiar phonological forms from long-term memory differ substantially from those involved in the retrieval of novel forms from short-term memory. A number of linguistic constraints appear to govern the types of error that occur in both tasks:

1. Many errors involve the substitution of one segment with another of the same size. Typically a single phoneme is replaced.
2. Consonants and vowels very rarely interact with one another; the CV-structure of the target is thus often preserved.
3. Substituting segments tend to be phonologically-similar along some dimensions to those they replace. Although the precise nature of this phonological similarity effect has not been clearly established, it appears that features defining with manner of articulation are more important than those defining place. Errors which breach the phonotactic constraints of the speaker's language are very rare.

## 4. Modelling constraints on phonological retrieval: a cyclical syllable template

---

The data reviewed in the previous chapter indicate that structural properties of the syllable constrain errors in the immediate recall of nonwords. Meanwhile, the general theories of serial order reviewed in chapter 2 have generally not been influenced by such linguistic considerations. The first section of this chapter shows how the involvement of syllable structure is crucial to the understanding of errors in phonological retrieval. We shall see that CQ models of the process which do not take account of linguistic structure produce errors which are not compatible with the data discussed in chapter 3. In section 2, a model of phonological retrieval in speech production (Dell, 1986) is briefly reviewed for comparison. Dell's is an implemented example of category of speech production models which provide an account of linguistic constraints on phonological errors. These theories involve representations which separate the structure and content of syllables. The structural representation

or 'frame' is used to constrain the order in which the phonemic content is output. The application of the same insights in a model of short-term memory might help to account for constraints on errors in immediate recall. However, models such as Dell's cannot explain how such representations are learned. A particular problem from the point of view of short-term memory is that the representation of syllable structure must be acquired in a single-trial. The third section develops a model of the process by which syllable structure can in principle be computed rapidly, on-line as the stimulus is perceived. This account of syllable structure will be central to the model of short-term memory presented in chapter 5.

### 1 Phonological slips, STM, and CQ models

It is clear from the data discussed in chapter 3 that spontaneous speech errors are structurally constrained. This is certainly the case with regard to phonological slips, which account for the bulk of speech errors. The same constraints operate in experimentally elicited slips. These data are consistent with the findings from experiments involving short-term memory, indicating that errors in the recall of nonwords tend to retain structural characteristics of the target items. In short, it seems likely that in a variety of experimental and real-life contexts, *including both spontaneous speech and short-term memory experiments involving nonwords*, errors in phonological sequencing could be accounted for by assuming that they are due to the flawed operation of the same mechanism responsible for ordering speech sounds. If this appears to be obvious then it is remarkable that models of short-term memory have never included speech-specific mechanisms for phonological retrieval. This omission is all the more surprising given the strong and independent evidence of the important role played by articulatory coding and control processes in verbal memory (outlined in chapter 1).

## 1. PHONOLOGICAL SLIPS, STM, AND CQ MODELS

---

To what extent can the mechanisms of phonological retrieval be reconciled with existing CQ accounts of serial order? To explore this question, this section discusses the ways in which CQ has been applied to phonological sequencing (Houghton, 1990; Glasspool, 1994).

The first application of the CQ architecture to phonological retrieval was by Houghton (1990). Houghton's model uses a very simple two dimensional control signal, as discussed in chapter 2. For any sequence to be learned, at the sequence level, a start node (initiator- or I-node) is active at the beginning. As the sequence is presented, activation of the I-node gradually decays. The E-node remains inactive, until the end of the sequence is experienced. A set of nodes representing phonemes to be sequenced are activated one at a time as they are experienced during learning, and weights between concurrently active sequence and phoneme nodes are strengthened at each timestep. The result is that phonemes at the beginning of the sequence are strongly associated with the I-node, and only weakly with the E-node. Conversely the phonemes at the end of the sequence are strongly associated with the E-node, and weakly associated with the I-node.

Recall proceeds by activating the I-node maximally. As it decays passively, the activation of the E-node at each timestep is the complement of the I-node's activation. This produces an activation gradient in the phoneme nodes, which is modulated during recall. In combination with the operation of a competitive filter, which serves to suppress a selected response at each timestep, this dynamic modulation of the input to the phoneme nodes, is sufficient to allow the correct recall of short sequences after a single presentation. Longer sequences required a supervised practice phase. To model phonological retrieval from short-term memory we can examine its performance after single trial learning. To do this, the model was implemented as set out in (Houghton, 1990). A layer of 47 phoneme (item) nodes <sup>1</sup> was connected to a

---

<sup>1</sup>For comparability, the same form of phonological coding was employed as is used in the model

## 1. PHONOLOGICAL SLIPS, STM, AND CQ MODELS

---

pair of word (sequence) nodes.

The first problem is that the Houghton model provides no mechanism for chunking continuous speech into manageable chunks. The two-dimensional control signal cannot learn sequences longer than about four items in a single presentation, so words containing more than a few phonemes cannot be recalled correctly. As the length (in phonemes) of the words increases, the performance of the model declines rapidly. Even in the absence of any noise in the system, words with more than four phonemes exceeded the model's single-trial learning capacity. Typically, after the initial phonemes had been recalled in the correct order, phonemes from the end of the sequence would be selected in advance of their intended position e.g., *graʊnd* → *graʊdn*, leading to (sometimes multiple) local transpositions of phonemes near to the end of the sequence. Such errors are largely due to the difference in the control signal experienced during learning and recall. During learning, the E-node is not active until the very end of the sequence, but during recall, the E-node becomes active much earlier on. This activates phonemes from the end of the list in advance of their intended position.

The model also shows evidence of difficulty in recalling shorter sequences containing repeated phonemes e.g., *pepə* ("pepper") → *peəp*, *bʌbl* ("bubble") → *bʌlb*. In such errors, the repeated phonemes can not recover from initial suppression in time to be produced in the correct position. This failure to recover from suppression could easily be overcome by altering the parameters of the model (so that for example the strength of feedback inhibition from the competitive filter was reduced).

For longer sequences, responses were usually extremely disorganised, and although the first few phonemes of any word were generally recalled in the correct order, the remainder of each response showed little structural resemblance to the in-

---

described in chapter 5.

---

## 1. PHONOLOGICAL SLIPS, STM, AND CQ MODELS

---

tended sequence, e.g., θəmɒmɪtə (“thermometer”) → θəmtɪəpt, juːnɪfɔːm (“uniform”) → juːnməʃi.

These difficulties in encoding long sequences are largely due to the limitations of Houghton’s two-dimensional control signal. A possible solution is provided by Burgess and Hitch (1992) who describe a control signal which is capable in principle of encoding sequences of arbitrary length without chunking. In their model errors resulted solely from noise and decay in the learned weights. The same scheme could be used for encoding sequences of phonemes, and this is the approach suggested by Glasspool (1994). In Glasspool’s model both words and nonwords are stored in this way (as sequences of phonemes). In addition words are also subject to lexical sequencing as used in the Burgess and Hitch model (i.e., as a sequence of words, each having a node corresponding to a prelearned lexical-phonological representation). These two sequencing systems interact during recall. Nonwords cannot be coded with the lexical sequencing mechanism as they have no pre-existing lexical representation. Nonword recall thus relies entirely on the phonological sequencing mechanism. According to Glasspool, this is essentially identical to the lexical sequencing system, except that the items to be ordered are phonemes rather than words.

The phonological sequencing system uses the same form of control signal as the lexical system – one adopting the modifications suggested by Burgess and Hitch. This has the advantage (over Houghton’s model) that long sequences of phonemes can be stored without chunking. The difficulty comes when one analyses the model’s errors qualitatively. In this regard, the important feature of the control signal is that it is temporally-autocorrelated. This, the reader will recall, is crucial in accounting for the characteristic pattern of mainly local transposition errors in standard serial recall tasks. Yet, as the data described in the previous chapter make clear, phonological ordering errors have a different character. One of the constraints on such

errors is that within-syllable (local) transpositions are very unlikely. Instead segments typically move to the same position in a nearby syllable. Moreover, it seems that consonants do not exchange with vowels or vice versa.

Although Glasspool's model provides a useful insight into the quantitative differences between word and nonword recall (e.g., it can account for the findings of Hulme et al., 1991), the errors it produces are qualitatively quite different from those which are observed empirically. David Glasspool (personal communication) has supplied simulation data which shows how the model's responses differ from those made by human subjects. The data is from a simulation of a typical nonword recall experiment (Treiman & Danis, 1988, experiment 1) using his own implementation of the model described in Glasspool (1994). In the experiment, subjects were required to recall lists of 6 CVC nonwords. In Glasspool's model, such sequences would be encoded as 'lists' of phonemes which are not differentiated in any way with regard to their structural properties (i.e., there is no distinction between vowels and consonants). After every third phoneme, a null phoneme (denoted  $\emptyset$ ) is encoded to mark the end of one nonword and the beginning of the next. A sequence like /ger, vaŋ, kus, dæɫ, jɔb, ʃim/, would thus be encoded as: /g/, /e/, /r/,  $\emptyset$ , /k/, . . . , /m/. Since all such lists would be encoded in the same way, the simulation involved the same six-item list being learned and recalled 1500 times. During recall, a syllable boundary marker competes with other phoneme nodes, and can transpose with them. A typical attempt at recalling the above target sequence is: /ger, vaŋuk, sd, æɫ, j, bɔʃim/.

Output is not necessarily organised into items which can readily be compared to the targets, for instance some of the model's responses contained more items than the target lists, others contained fewer. It is therefore not always clear which response relates to which target. The model in fact made more responses (9689) than there were target syllables (9000). Responses were marked as correct if they were in

## 1. PHONOLOGICAL SLIPS, STM, AND CQ MODELS

---

the right serial position and had the correct phonological form. Only the first six responses to any list were marked in this way. Of these, 2660 were correct, leaving 7029 errors of commission. The CV structure of each response was calculated; these results are shown in table 4.1. The number of phonemes in each response was also counted. The proportion of erroneous responses at each length is shown in table 4.2.

Though the supplied data are from a simulation with a rather high error rate compared with humans (29.56% of responses correct in the simulation, 43% in Treiman and Danis' experiment), the analysis shows a fundamental qualitative incompatibility with the empirical observations. The absence of any representation of syllable structure means that Glasspool's model produces a relatively small proportion of responses which retain the syllable structure of the target items (see table 4.1), and most of these are correct responses. For example, while Treiman and Danis found that 96% of errors had the same CVC structure as the target items, only 794 (11.30%) of the erroneous responses produced by Glasspool's model had this structure. It also produced a high proportion of very implausible responses which could not be pronounced, for instance, 11.25% of all responses were singleton consonants. In fact, because syllable boundary markers can transpose with phonemes, the model produces many responses which drastically change the length (in phonemes) of an item. Such responses are not observed experimentally: table 4.2 shows a comparison of the number of phonemes in the model's erroneous responses and those of adult subjects in a similar experiment (Treiman, in press). A further consequence of the lack of a structural representation of the syllable is that within-syllable errors such as /ger/ → /egr/ can readily occur. Such errors are not infrequent in the model's output, yet the most common errors (Treiman & Danis, 1988), in which a phoneme moves from one syllable to an analogous position in another syllable are rare. It is also often the case in the model's errors that a syllable initial phoneme moves to a syllable final position or *vice versa* (e.g., /ger, vŋ/... → /gerv, ŋ/...).



# 1. PHONOLOGICAL SLIPS, STM, AND CQ MODELS

---

Table 4.1: *The proportion of responses showing various consonant-vowel structures in simulations of serial recall of lists of six monosyllabic nonwords.*

CV-structure	% responses
CVC	32.55
C	11.25
CV	10.57
VC	7.90
VCC	5.84
CCV	5.83
CCVC	4.87
CC	3.83
CVCC	3.42
V	2.03
other	11.91

Table 4.2: *The proportion of errors produced by Glasspool's model with different phonemic lengths. Data from human subjects performing a similar task are shown for comparison (Treiman, in press)*

Error Type	% of total errors	
	Adults	Simulation
Omission	10.9	—
1 phoneme responses	0.0	19.33
2 phoneme responses	2.0	31.47
3 phoneme responses	83.7	25.62
4 phoneme responses	3.2	15.78
5 phoneme responses	0.1	4.34
6 or more phoneme responses	0.0	3.46

The failure of CQ models to predict the characteristic pattern of errors in nonword recall is due to two related factors: first, no existing CQ model provides a plausible mechanism for the chunking of speech into syllabic units, instead they assume that phonemes are ordered in the same unstructured manner as any other action sequence might be. Second, CQ models generally assume a temporally-correlated control signal which tends to produce local transposition errors; the very type of errors which are prohibited in phonological retrieval by structural constraints. The latter property, though, is useful in explaining errors in serial recall in situations where there are no such constraints. Therefore, the approach I will take in this thesis is to develop the basic competitive queuing model of short-term memory, incorporating an additional mechanism for the structural representation of serial order at the phonological level. The remainder of this chapter explores the issue of how a structural representation of the syllable might interact with serial ordering mechanisms. This issue has not previously been considered in general models of sequencing<sup>2</sup> (those considered in chapter 2). However, in chapter 3, error data from spontaneous speech production were found to show a pattern of phonological constraints consistent with those seen in nonword recall and it was suggested that the same mechanism was involved in phonological retrieval whether from short- or long-term memory. Therefore, consideration of psycholinguistic accounts of phonological retrieval in spontaneous speech is likely to shed some light upon the issue.

## 2 Psycholinguistic models of phonological retrieval

The primary aim of linguistic grammars is to describe the productivity of language as concisely as possible, and not to explain speech errors. Such analyses are based

---

<sup>2</sup>One exception is Shaffer's (1976) model which provides an interesting but incomplete account of the way in which apparently similar linguistic factors might be involved in typing

---

## 2. PSYCHOLINGUISTIC MODELS OF PHONOLOGICAL RETRIEVAL

---

in the main upon distributional data. It is thus remarkable that (as noted in the previous chapter) the interacting units in speech errors generally correspond to the units of linguistic grammars. This can be taken as evidence that linguistic grammars are not merely abstract structures for the description of language, they also lie at the heart of the human speech production system. It is thus quite natural for linguistic models of speech production to start from the premise that speech is represented in the brain in terms of the hierarchical structures of generative grammars. Theories based on such structures share the assumption that producing an utterance involves assigning values to variables at various levels of a tree-like structure. This variable binding process can be seen as a kind of frame-and-slot mechanism. First of all a frame is specified. The frame specifies the order of a set of categories. For example, producing a sentence might involve first selecting a frame e.g.,  $N_{pl}VN_{pl}$ . Here, each symbol corresponds to a category of word which can fill a 'slot' in the frame ( $V$  for a verb, and  $N_{pl}$  for a plural noun). The slots are filled with items which are strictly constrained by the category (e.g., "cats eat mice"). Which items fill the slots is determined by the semantic system. Each frame can be filled by many different items ("people prefer peas", "whales like plankton"), but the frame serves to constrain serial order. Some (non-grammatical) orderings are prohibited ("people peas prefer"). However there is scope for errors to occur. Firstly, the wrong item can be inserted into a slot ("cats eat dogs"). Secondly, there may be some ambiguity about the order in which the items are to be inserted into the frame – in this instance, the two nouns can be exchanged without violating the frame constraints: "mice eat cats".

Similar frame and slot mechanisms are proposed for different stages of the production system. Once the words have been selected it is necessary to specify their phonological forms. Like syntactic frames, the nature of phonological frames has been elaborated by analysing distributional properties of languages – in any language some phonological forms are possible, whereas others never occur (e.g., the

## 2. PSYCHOLINGUISTIC MODELS OF PHONOLOGICAL RETRIEVAL

---

cluster /tɫ/ cannot precede the vowel in any English syllable – I will return to such analyses later). A phonological frame might take the form *onset*, *peak*, *coda*, where each stands for a category of phoneme or phoneme cluster which can fill a particular position within the syllable. Note that this prohibits many local transpositions between phonemes, since the phonemes which can fill one slot will not generally overlap with those which can fill another.

Models with this “frame and slot” character have been very successful in accounting for a wide variety of errors at various levels of linguistic analysis (e.g., Shattuck-Hufnagel, 1979; Mackay, 1972). Until quite recently, though, models of speech production were fairly informal, lacking sufficient specification, in particular of the underlying serial ordering mechanism, to be truly testable. In the late eighties, though, the situation was improved with the advent of models which combined ‘spreading activation’ (connectionist) architectures, responsible for specifying structure and content, with implicitly competitive output processes, responsible for serial order. One of the most widely known is Dell’s (1986) model of sentence production, which was concerned with errors at the phonological level. The Dell (1986) model proposed that the phonological content of a planned utterance is determined by the spread of activation in a hierarchical network. At the top of the hierarchy are units representing the words and syllables to be articulated. Lower down are units standing for the phonemes and the consonant clusters which are to be articulated. Separate groups of units are used to represent segments from syllable onsets, peaks and codas, so that the /k/ in “cat” would be represented by a different unit (designated  $k_{onset}$ ) from the /k/ in tack (designated  $k_{coda}$ ).

Planning involves activating units at the lexical/syllable level, and allowing activation to spread downwards via the excitatory links which exist between lexical and phonological units. Lexical units are activated in parallel: the sooner a word is to be articulated, the more active is the unit which represents it (cf., Burgess &

## 2. PSYCHOLINGUISTIC MODELS OF PHONOLOGICAL RETRIEVAL

---

Hitch, 1992). The parallel activation of the word nodes leads to parallel activation in the phoneme nodes. When articulating a phrase like “reading list”, the unit representing the  $l_{onset}$  of “list” would become active while “reading” was still being planned. The control process responsible for ordering the sounds for serial articulation is competitive: for each syllable, the most active onset, peak and coda units are selected in that order. For example, “read” would be articulated by selecting first  $r_{onset}$ , then  $i_{peak}$ , and finally  $d_{coda}$ . The onset, peak, and coda correspond to slots in a syllable frame. Errors such as “reading list → leading list” could be explained by suggesting that the activation of the intruding segment ( $l_{onset}$ ) exceeded that of the target segment ( $r_{onset}$ ) when a particular slot (here, the syllable onset) is being filled. It was proposed that this might occur given a limited amount of ‘linguistic noise’ in the system, perhaps spreading from many partially-activated lexical units, activated in parallel by the semantic system. Simulations using a simplified form of the network showed a similar distribution of errors to those observed in collections of naturally occurring speech errors. Although Dell’s (1986) model followed a long line of theories which explained error data in terms of the interaction between linguistic structure and phonological content, his was one of the first accounts which was sufficiently well-specified to be tested against error data. The features of the model that made this possible were its assumption of parallel activation in the planning of speech, and the use of a competitive mechanism to determine which segments filled each slot in a syllabic frame. Linguistic structure is represented explicitly both in the structure of the network, and in the nature of the serial control process which is used to read-out the parallel structure.

Dell (1988) adapted the model to use wordshape (CV-structure) representations to provide a mechanism for constraining the order of the phonemes in the output. The structural (wordshape) information is represented separately from the content (phonological) information. The two representations interact in the production process. Position-specific coding was not abandoned, so a syllable-initial phoneme was

## 2. PSYCHOLINGUISTIC MODELS OF PHONOLOGICAL RETRIEVAL

---

still represented by a different node to a syllable-final one. Dell (1988) provides a hint as to how a phoneme-level sequencing mechanism might be implemented at the level of the individual nodes:

“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.”

Thus Dell's model 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, Dell restricts response competition at any syllabic position to those phonemes which can occupy that position (according to a syllabic template). In the first formulation of the model, one general syllable template was used in the production of each syllable, in the second version, each syllable was associated with one of a set of possible templates.

In both versions of the model (Dell, 1986, 1988), 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). Dell's model thus shows how the linguistic structure of the syllable might constrain the serial order of segments in a model which is broadly compatible with the CQ framework. However, the model was directed at errors of retrieval in spontaneous speech production (from a pre-existing long-term phonological lexicon). The problem of how serial information might be learned was not addressed. In addition, the competitive mechanism responsible for ensuring serial output was only rather sketchily described. The use of position-specific coding was (as Dell conceded) undesirable, and yet central to the sequencing mechanism

### 3. GENERAL CONSTRAINTS ON SYLLABIC PHONOLOGY

---

proposed. However the basic concept of separating content from structure (which is used for sequencing) provides an appealing account of production errors.

In order to form a 'framelike' representation of a new syllable in a single trial, its structural properties must be abstracted while it is being perceived. The next sections explore the way in which this is achieved, an issue which has not previously been considered in any model of phonological memory.

### 3 General constraints on syllabic phonology

In considering phonological constraints at and below the level of the syllable, it is important to concentrate on those which apply cross-linguistically. The most important of these is the sonority principle. Sonority is a linguistic quality associated with each segment, which is related to the energy of the sound it represents. The sonority principle (see e.g. Selkirk, 1984) imposes constraints upon the order in which segments can occur in a syllable. For any language, each phoneme can be assigned relative sonority values, specifying its place in the hierarchy (see figure 4.1a). The relationship between the sonorities of successive segments in a well formed syllable appears to conform to a very general pattern; successive elements tend to increase in sonority towards a single peak (usually a vowel), and then decline (figure 4.1b). The relationship between specific phonemes is controversial but it is possible to make generalisations about categories of phonemes: obstruents have low sonority; nasals and liquids have intermediate values, while vowels are the most sonorous segments. The principle can be used to determine whether or not a sequence of phonemes forms a single well-formed syllable or not. For instance, sequence /flʌnt/ conforms to the generalisation whereas sequences like /lfʌnt/ and /flʌtn/ do not.

The constraints imposed by sonority are supplemented for a particular language

### 3. GENERAL CONSTRAINTS ON SYLLABIC PHONOLOGY

---

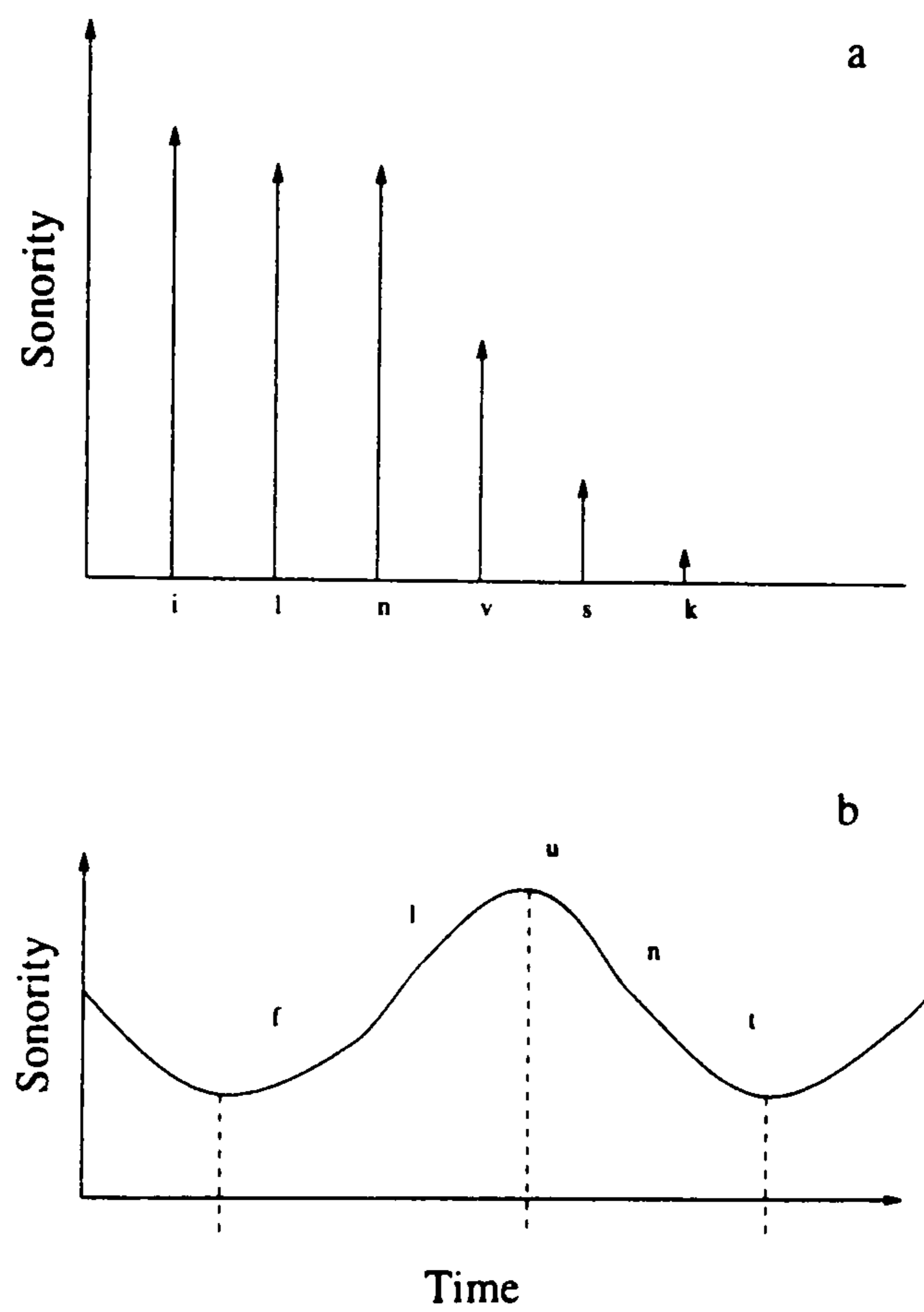


Figure 4.1: *The sonority principle: a) each phonological segment is associated with a particular sonority value, reflecting its energy relative to other segments (based on a diagram from Ladefoged, 1993); b) in a well-formed syllable, segments are usually organised such that sonority rises to a single vocalic peak, and declines thereafter.*



### 3. GENERAL CONSTRAINTS ON SYLLABIC PHONOLOGY

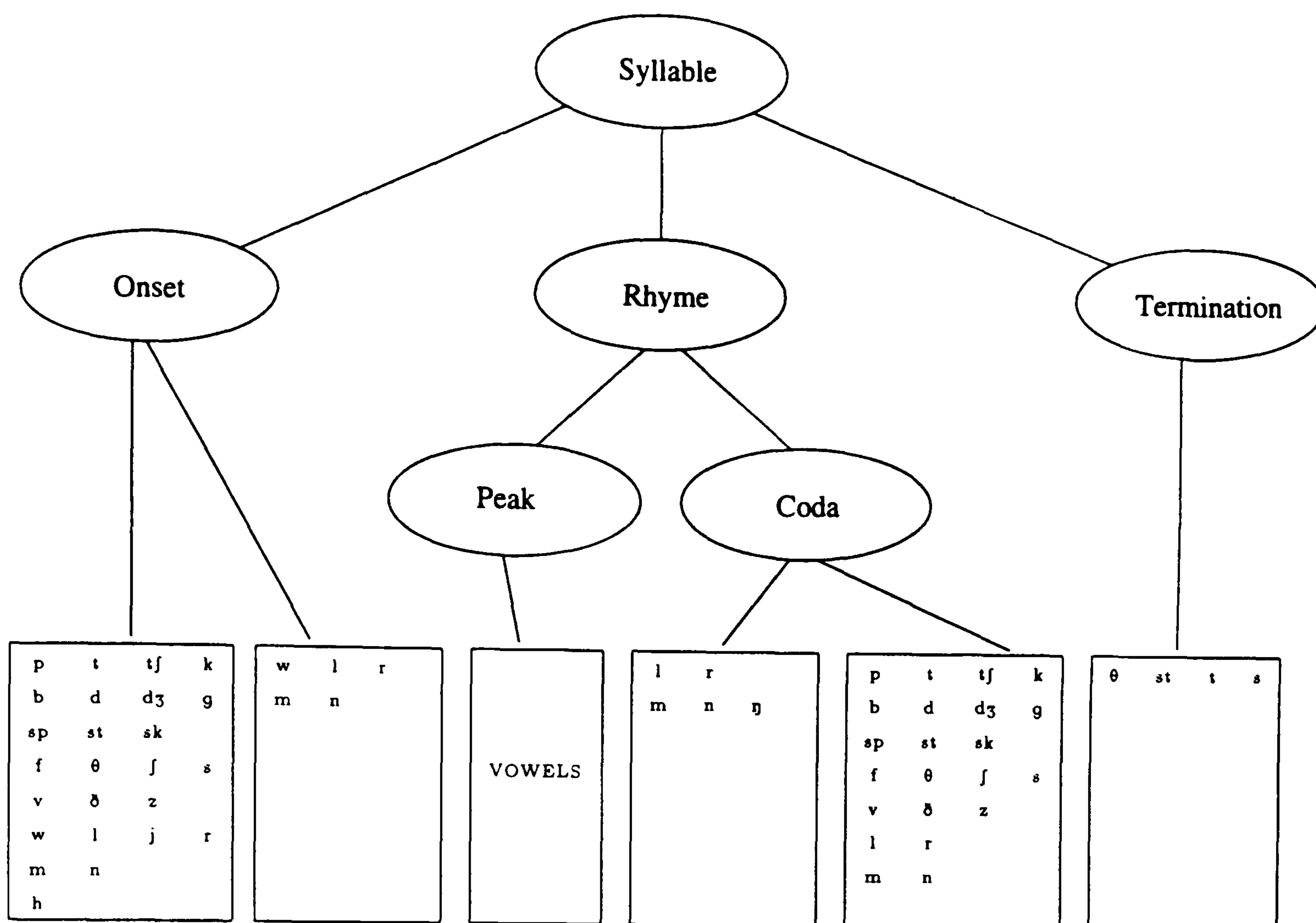


Figure 4.2: *English syllable structure as described by Fudge (1969)*

by further 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 captured in the template described here, and represented in the model described in chapter 5.

Fudge (1969) set forth a number of rules to provide a productive description of legal English syllables. The central structure underlying the grammar is summarised in figure 4.2. The syllable is seen as a hierarchical entity comprising onset, rhyme and termination units. Each of these sub-syllabic units (sometimes referred to as syllabic constituents) is comprised of a number of slots, each of which can be filled by a subset of phonemes, or remain empty, for any well formed syllable. Note that the clusters /sp/, /st/, and /sk/ can each occupy a single-slot, the remaining basic elements of the grammar are phonemes.

With the exception of the termination sub-unit, which applies only to word-final syllables, there is a noticeable symmetry between the phonemes that can fill pre- and post-vocalic slots; this is a reflection of the operation of the sonority principle. The slots neighbouring the vocalic peak (always a vowel in English) can only be filled by nasals, liquids, and glides, the most sonorant consonants. The outer slots can contain the less sonorant obstruent segments as well.

### 4 A syllabic template

The slot positions described by Fudge have been used as the basis of a general syllable template, which can be used to represent legal syllable structures. It is important to note that this account of syllable structure is based solely upon distributional evidence. The structure proposed here thus has a theoretical motivation independent of any consideration of the error data (chapter 3) which are to be modelled. The template adopted in the current model is shown in figure 4.3. For reasons that will become apparent it is depicted as a cyclical structure. The template comprises 5-slots representing the five non-termination positions in Fudge's syllable structure. Each slot can be 'filled' by a subset of phonemes listed next to it. 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.

#### 4.1 Parsing the syllable

The constraints arising from the sonority sequencing generalisation make it possible to describe the structure of most legal syllables in a single cycle by progressing in clockwise direction to the next matching slot for each phoneme encountered in a sequence. The sequence /flʌnt/ involves the slots 1,2,3,4 and 5; pat involves the

#### 4. A SYLLABIC TEMPLATE

---

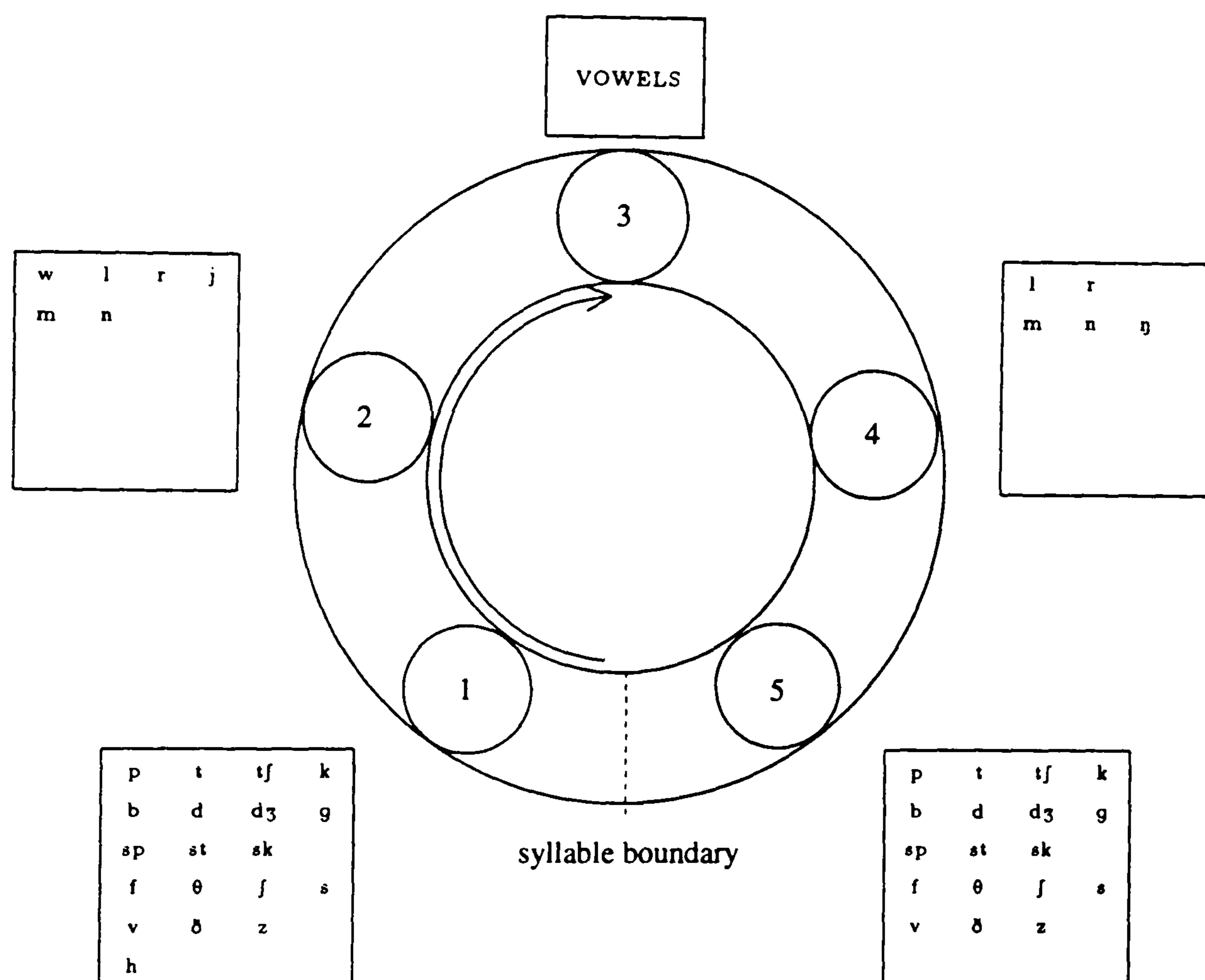


Figure 4.3: *The structure cyclical syllabic template used in the model described in chapter 5*

#### 4. A SYLLABIC TEMPLATE

---

slots 1,3 and 5. A large number of sequences that are not legal syllables would entail more than one cycle e.g. /lfʌnt/. A few legal (English) syllables cannot be ‘parsed’ in a single cycle, for example /bedz/. Such syllables involve either the use of a termination position, or contain two successive phonemes from the same slot. A number of phonotactically illegal (in English) syllables can also be parsed in a single cycle, for example /tɒrk/. The syllable template is clearly something of a simplification, and as such it fails to capture all the regularities of English. However it is sufficient to capture the general constraints arising from the sonority principle, and to represent most legal syllables.

In the version of the model used in the simulations described in chapter 5 the cyclical activation of positions in the template is implemented by rule. However, it is possible to specify a neural network capable of performing the same function. This network is described in appendix D.

#### 4.2 Parsing continuous ‘speech’

As well as representing the internal structure of the syllable, the cyclical nature of the template makes it possible to parse a continuous phonological input stream into syllables. Whenever a new cycle begins, a new syllable has been encountered. Table 4.3 shows how the continuous sequence /rʊmelhɑtændnɔmən/ (“Rumelhart and Norman”) is parsed into syllable-sized chunks.

As I pointed out above, the template proposed here is not intended to capture of all the phonological constraints of a particular language; nonetheless, it is possible to test its effectiveness by comparing its performance with standard syllabifications specified in a psycholinguistic database. The corpus used contained all of the 3177 words in the Oxford Psycholinguistic Database which had both a Thorndike-Lorge frequency greater than 1000 and which were phonetically transcribed with the po-

#### 4. A SYLLABIC TEMPLATE

---

Table 4.3: *A continuous stream of phonemes is parsed into syllable-sized chunks by assigning each phoneme to the next available clockwise slot.*

<u>Sequence</u>	<u>Slots used</u>
/rʊm/	2,3,4
/el/	3,4
/hæt/	1,3,5
/ænd/	3,4,5
/nɔm/	2,3,4
/ən/	3,4

sition of syllable boundaries marked. The syllable template was used to parse each phonetically encoded word into syllables, with a syllable boundary marked each time a new cycle of the template began. The results were compared with the standard segmentations as coded in the database. The template's performance was found to be remarkably good for such a simple mechanism. The number of syllables identified using the template agreed with the psycholinguistic database for 3083 (97%) words. The exceptions are listed in appendix C. In most cases they are words which involve either two consonants from the same slot (e.g., act) or demand the use of the termination position which has been omitted from the template (e.g., box).

For 1484 (47%) of the words the position of syllable boundaries specified using the sonority wheel template were identical to those in the database. In most of the remainder, the cyclical template assigned a phoneme to the rhyme of one syllable, where standard syllabification rules suggest it should be assigned to the onset of the next. For instance, the template suggests that "agony" should be parsed as /æɡ—ən—i/ (where the syllable boundary is represented by —). In the database it is coded as /æ—gə—ni/. The principle involved is called the maximal onset rule, which states that where an intervocalic consonant can be assigned to either of two syllables, it should be assigned to the onset of the second syllable if it is

#### 4. A SYLLABIC TEMPLATE

---

phonotactically legal to do so. It is important to note, however, that the maximal onset rule does not rigidly constrain subjects' responses in experiments where they are asked to make judgements concerning syllable boundaries. Such experiments show that there is often some ambiguity about the status of intervocalic consonants (Treiman, Straub, & Lavery, 1994). Therefore the differences in the template's assignment of intervocalic consonants is not too serious a problem. The benefit of not implementing a maximal onset rule is that each phoneme can be parsed 'on-line' without looking ahead at the rest of the sequence to determine whether or not it would be legal in the next syllable's onset. This simplification is necessary to properly model short-term memory tasks where presumably such look-ahead is impossible. The simple version of the template described above is quite adequate to perform the function of representing syllable structure in short-term memory, and works within the constraints imposed short-term memory tasks. Issues such as the syllabification of intervocalic consonants would be better explored using either of the more detailed, low-level parsers described in appendix D, but such issues lie beyond the scope of the current work.

# 5. A new model of short-term memory

---

In this chapter I present a new connectionist model of short-term auditory-verbal memory. In this account, the linguistic constraints on syllabic phonology are made explicit in the system responsible for the rapid learning, temporary storage and retrieval of serially-ordered phonological information.

## 1 Description of the model

The primary aim of the model is to incorporate phonological structure into verbal short-term memory. To do so, the model must simultaneously represent item and order information at (at least) two levels, and be able to construct these representations in a single exposure to a stimulus list. The two levels at which the current model operates are (i) the syllable level, and (ii) 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, /ger, vɑŋ, kus, dæɪ, jɔb, ʃɪm/. Errors might involve the reordering of entire syllables, e.g., /ger, vɑŋ, dæɪ, kus, jɔb, ʃɪm/. Phoneme-level errors could also occur, recombining phonemes from different syllables, or introducing phonemes not present in the target syllables e.g., /ger, vɑs, dul, kæm, jɔp, ʃɪm/.

In the model described below, 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 syllabic structure component, which tracks the input stream in real time, parsing it into syllables, and generating a representation of syllabic structure and content able to support immediate repetition of novel stimuli.

Given the hierarchical nature of the task being modelled, it is inevitable that the detailed dynamics of the resulting model will be somewhat complex, and may be difficult to grasp from a “static” description. This chapter sets out to provide a clear and intuitive account of the model’s operation. A full mathematical description of the model is provided in appendix B, where such technical issues as calibration of the model and the treatment of time and duration are also discussed.

### 1.1 Architecture.

The model is implemented as a neural network model, in line with relevant previous work (Burgess & Hitch, 1992; Dell, 1986; Houghton, 1990). The overall architecture of the model is shown in figure 5.1. Figure 5.2 is a more detailed diagram, showing



## 1. DESCRIPTION OF THE MODEL

---

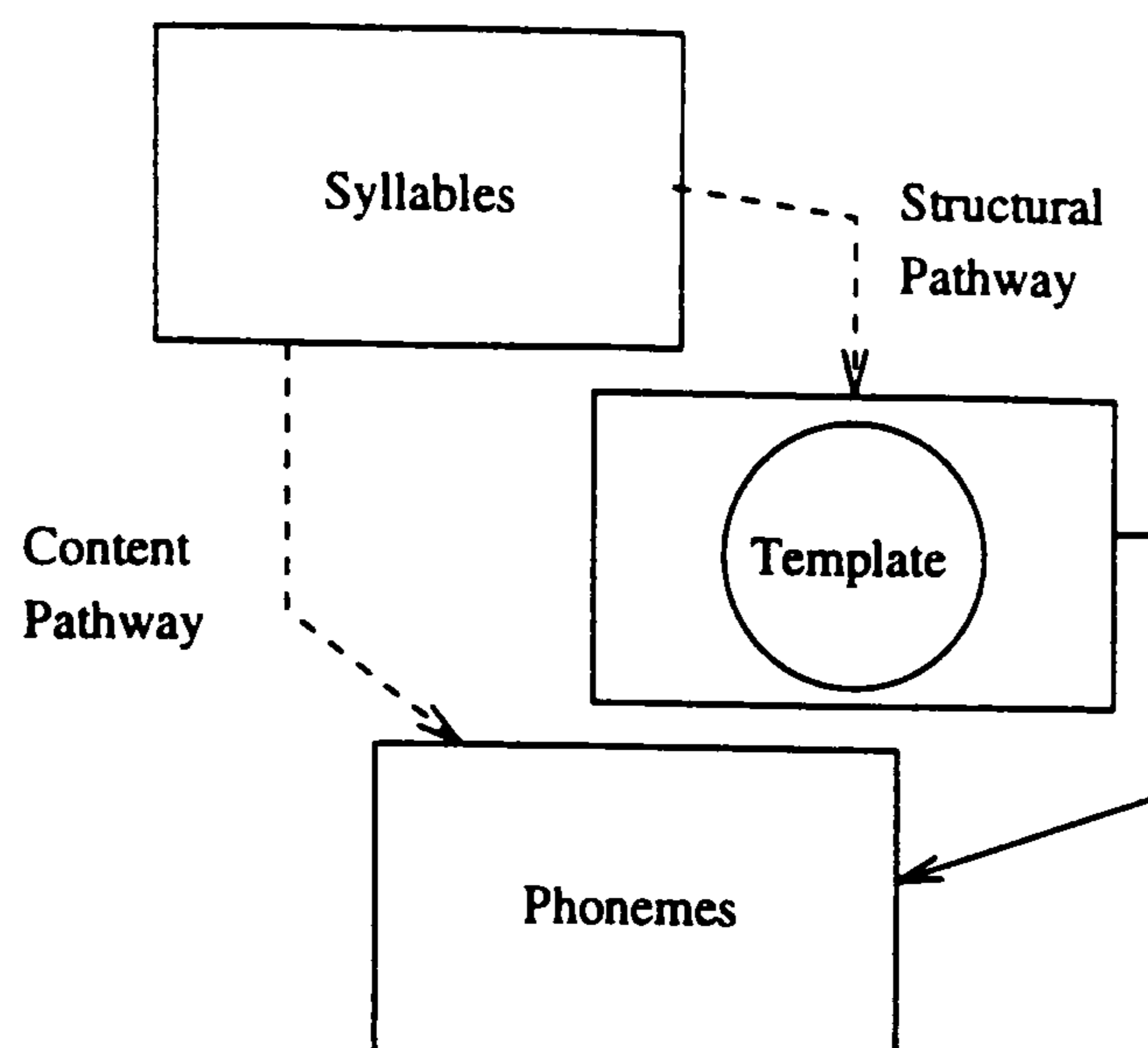


Figure 5.1: *An outline of the model's structure*

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 to units representing their constituent phonemes. One pathway (the content pathway) directly links the syllable unit to its constituent phonemes, the other (the structural pathway) operates via a general syllable template. Hard-wired excitatory connections link the 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, the value of which is set during learning (described below). Following Burgess and Hitch (1992) it is postulated that these connections used for short-term storage are temporary – their strengths decay over time to prevent saturation of the system.

The phoneme, syllabic template and syllable groups will now be described in more detail.

*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 (fol-

## 1. DESCRIPTION OF THE MODEL

---

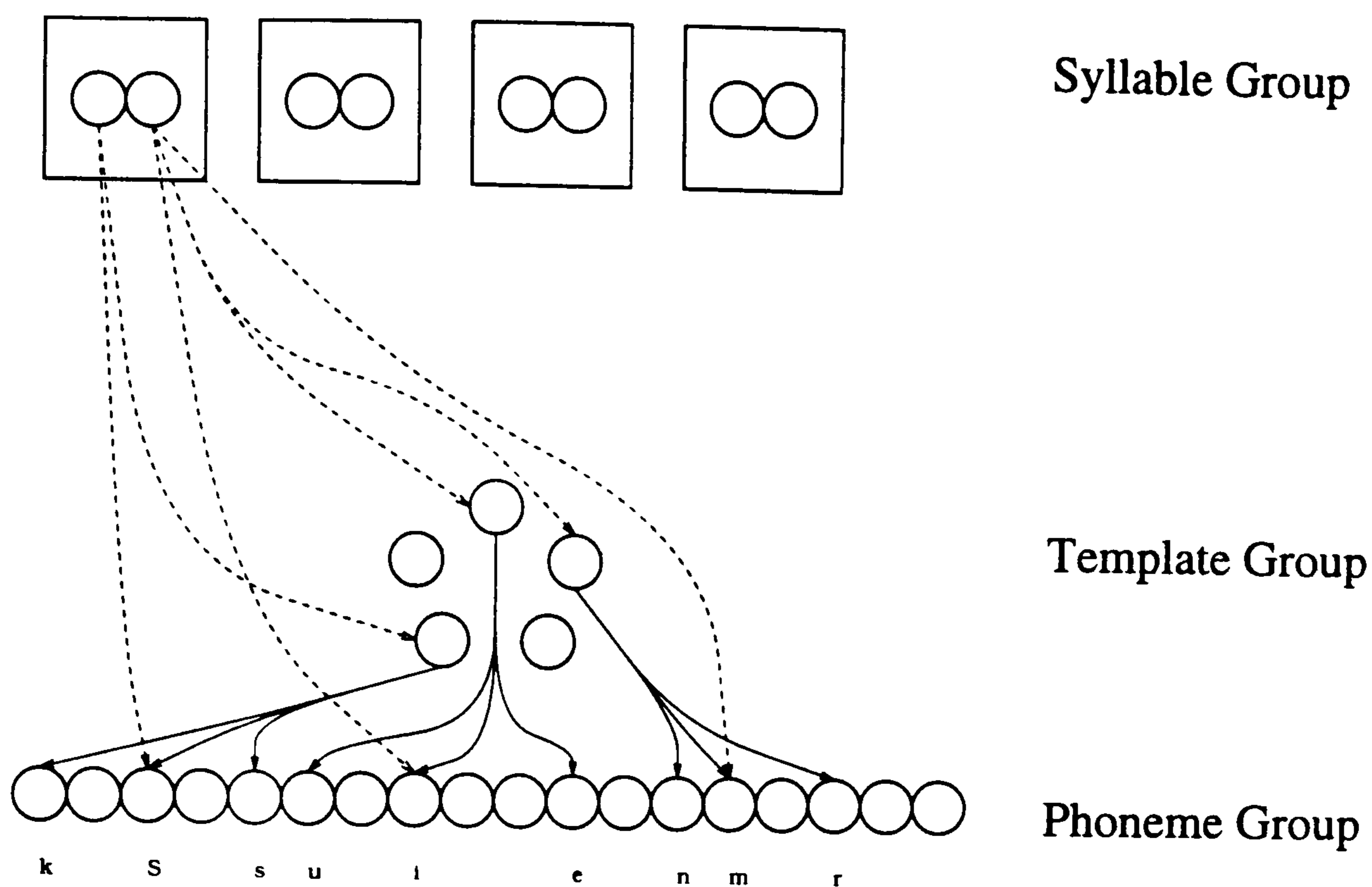


Figure 5.2: A more detailed diagram of the model's structure showing nodes and connections. The dashed lines represent temporary weights, the solid lines permanent connections.

lowing Fudge, 1968, see chapter 4).

*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 4.2). 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

## 1. DESCRIPTION OF THE MODEL

---

onset/rhyme division is necessary to distinguish such syllables from each other<sup>1</sup>; 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”, as described in chapter 4. The template is theoretically-motivated by linguistic accounts of phonological retrieval in speech production which suggest that a representation of the structural properties of an utterance are used to constrain the order in which the phonemes within each syllable are produced. At the phonological level, ordering processes are thus broadly consistent with linguistic models of phonological retrieval.

## 1.2 Learning and Recall

Running the model in simulations consists 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. These phases are now described in

---

<sup>1</sup>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. For this reason, position specific coding has been avoided

more detail.

### 1.2.1 Learning

During learning, the network is presented with streams of phonemes which it parses into 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. Below, the changing pattern of activation during learning is briefly described for each group 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 slot 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. It is therefore assumed 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

## 1. DESCRIPTION OF THE MODEL

---

available clockwise slot which can be filled by that phoneme. Consonants following the vowel will therefore activate post-vocalic consonant slots, if possible. A detailed neural network mechanism which shows this behaviour is described in appendix D.

As nodes in the various groups become activated in this way, the weights on the connections between them are altered. Weights are changed according to a Hebbian learning rule, such that concurrently active nodes have their (temporary) connections increased in strength. In addition, the strengths of the temporary connections also passively decay towards zero. For any two nodes (designated  $u_i$  and  $u_j$ ) having activation values of  $a_i$  and  $a_j$  respectively, and linked by a temporary connection of weight  $w_{ij}$ , the change in the weight at each timestep ( $\Delta w_{ij}$ ) is given by:

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

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

In some of the simulations described below, 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  $d$ ),  $\delta$  is given by:

$$\delta = 0.5^{\frac{d}{h}} \quad (2)$$

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). This process is illustrated in figure 5.3.

# 1. DESCRIPTION OF THE MODEL

---

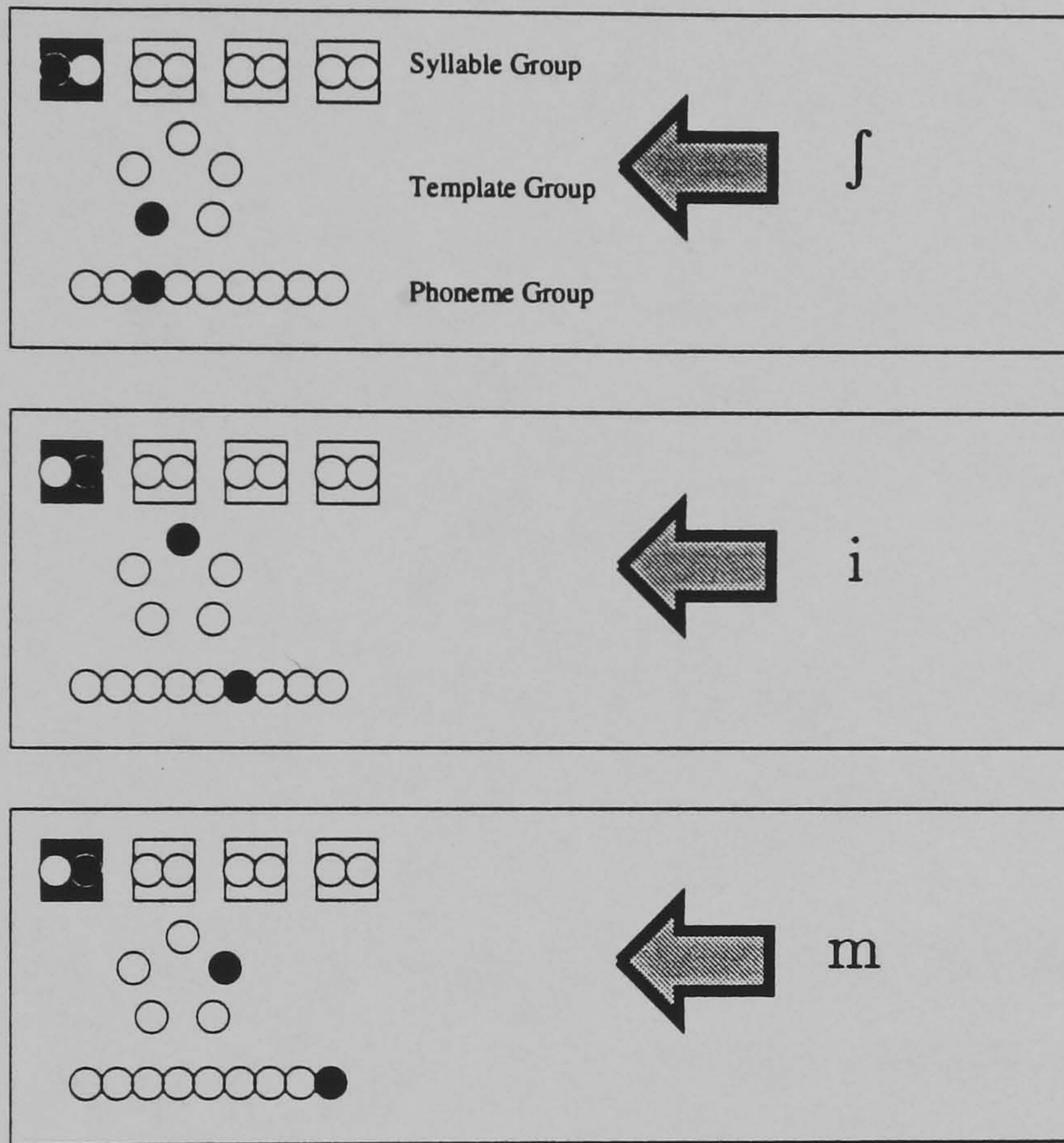


Figure 5.3: *During learning, the changing state of the input stream (represented by phoneme nodes) determines the activation state of nodes in the syllable and template groups. This diagram shows how the state of the network changes during presentation of a monosyllabic nonword. The activation of each 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.*

### 1.3 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 equation 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 figure 5.4.

*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

1. DESCRIPTION OF THE MODEL

---

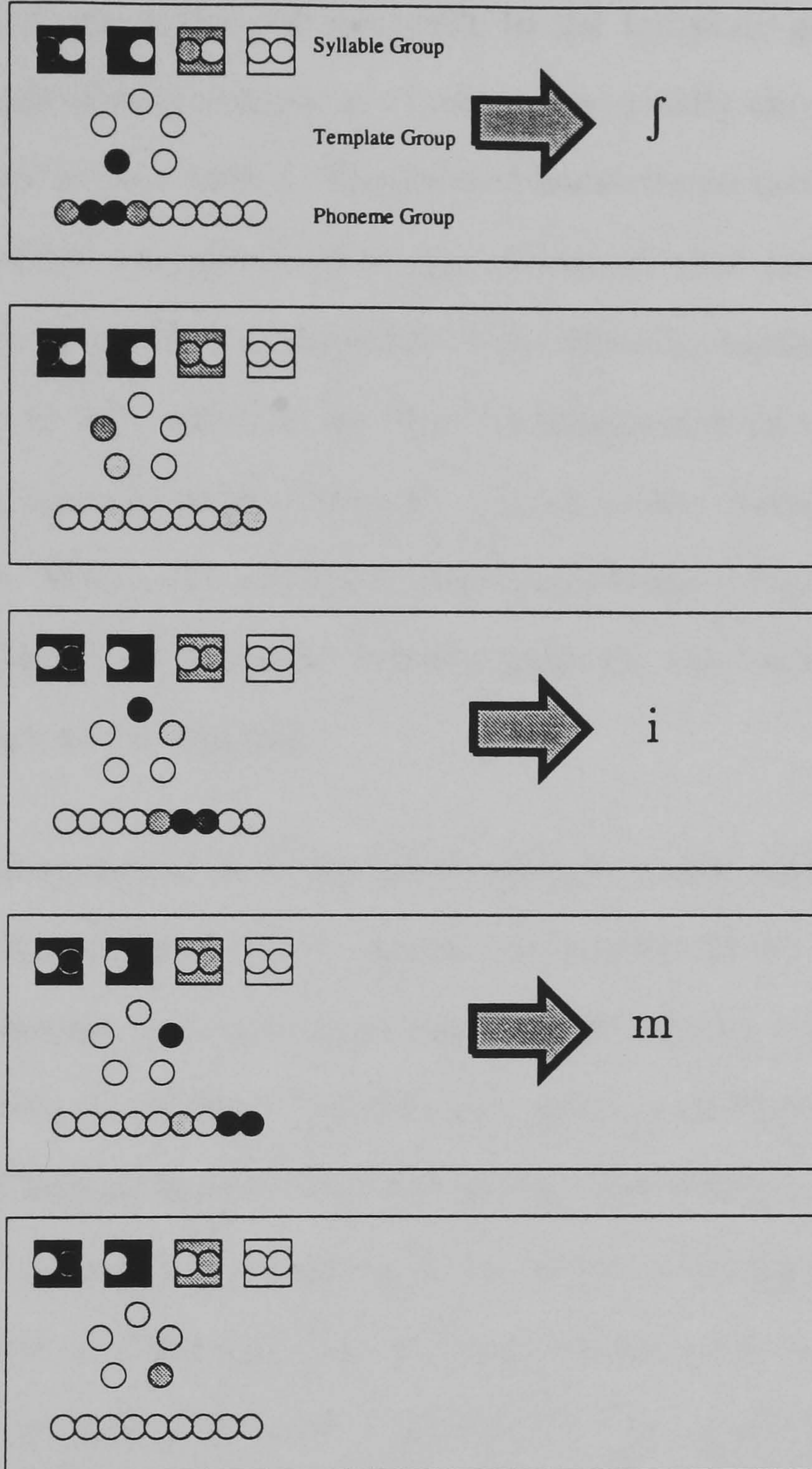


Figure 5.4: Diagram illustrating activation states of the network during recall of a nonword. As in figure 5.3 the degree of activation is represented by the density of the shading.



## 1. DESCRIPTION OF THE MODEL

---

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. It is proposed that 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 ( $\epsilon$ , 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'<sup>2</sup>) 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 B for further discussion). However, a phoneme node will become active if it receives combined input from both structural and content pathways. The model thus implements a scheme for dynamic variable binding. The variables in question are the syllabic slots and the bindings are phonemic values. 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

---

<sup>2</sup>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.

## 1. DESCRIPTION OF THE MODEL

---

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. Variable binding is therefore a competitive process.

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. This is because the level of input from the syllable nodes declines due to weight decay, so that for longer lists there is a good chance that, rather than the correct phoneme being selected, one of its competitors will be output instead. Often the substituting phoneme will be one from the same slot in a forthcoming syllable in the list. Occasionally, and especially when a good deal of time has elapsed between learning and recall, a phoneme from outside the list (but within the same syllabic slot) will gain enough activation to 'beat' the target phoneme and be selected for output. Error mechanisms at the root of these and other error types (insertions and deletions) are discussed in more detail in the section on simulations.

The performance of the model at recall is 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.

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 described below steps were taken to ensure that, so far as possible,

the time between presentation and recall of each phoneme is the same for all the phonemes in any sequence to be learned. The way in which this is achieved is described in appendix B. The rate at which syllables are generated at recall will also affect the performance of the model. 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.

## 2 Simulations

This section describes the procedure and results from simulations of experiments involving the brief retention of unfamiliar stimuli, and their serial recall. The mechanisms underlying some of the common forms of phonological error are then described in some detail.

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* to the effects modelled here, errors of the syllable sequencing system would also be expected. This sequencing mechanism is assumed to be like the one described by Burgess and Hitch (1992) for the ordering of familiar words which 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 towards 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.

The first section below explores quantitative aspects of the model's performance. In the second, qualitative breakdowns of the errors the model produces are presented.

## 2. SIMULATIONS

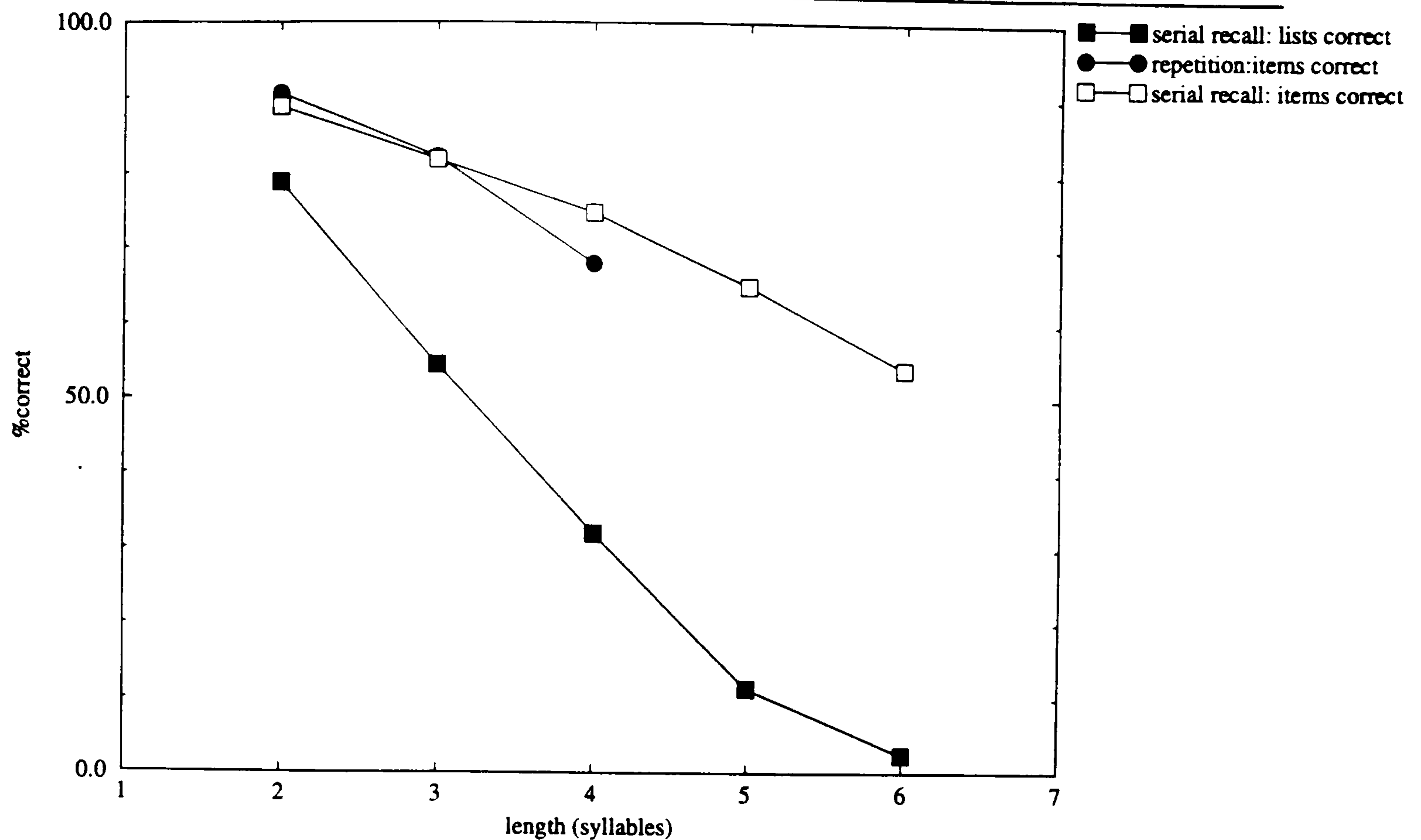


Figure 5.5: Graph showing proportions of correct responses in simulations of non-word repetition for items of varying length in syllables (black circles). Each point is plotted from 500 presentations. The results of nonword serial recall simulations are plotted on the same axes: both individual items (white squares) and lists (black squares) were scored for lists of varying length. Each point is plotted from 1500 list presentations.

The model's performance is compared with available empirical data.

### 2.1 Quantitative measures of performance

The following simulations show how the model's performance changes for lists of varying length, and how the serial position of a stimulus item affects the likelihood of its being correctly recalled. As noted above, these effects are expected to occur in addition to those resulting from errors at the syllable level.

So far as list length effects are concerned, if a hierarchical model of sequencing is to be adopted, then most errors must be at the phonemic sequencing level. Span for

## 2. SIMULATIONS

---

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 5.5 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/, /pɜplɪstərɒnk/). 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 circles 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 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,

## 2. SIMULATIONS

---

a longer sequence also demands that more phonemes are selected, and so there is a 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 errors at the syllable level also have an impact on span, the figure is a little on the low side.

Currently, I am aware of no data from adults which can be compared directly with the model's performance on the nonword repetition task. Most studies using the task have used either children or neuropsychological patients as subjects. Developmental changes in ability on the nonword repetition task can be modelled by varying the rate of weight decay in the network, and it is hypothesized that rapid rates of decay is characteristic of young children, and of patients suffering certain neuropsychological deficits (see section 2.3, "Modelling a damaged/developing store"). Like the results from the span simulations, the nonword repetition results show performance nearing, but slightly below what might be expected of normal adults (it is comparable with that of an 8 year old child). This suggests that the rate of decay in the model was a little too fast.

### 2.2 Qualitative analyses

In relation to the previous work in the CQ framework, the most important innovation in the current model is the use of a separate dynamic representation of syllable structure. This information can be used to supplement a representation of content

## 2. SIMULATIONS

---

at recall, ordering the segments, and limiting the space of potential errors. 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 the qualitative features of the model's output are analyzed and compared to analyses of responses obtained from human subjects.

The most detailed empirical data comes from Treiman and Danis (1988). This experiment was simulated using the same 30 lists of 6 monosyllabic CVC nonwords they used (table 5.1 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), hence the error rates are reasonably comparable. The responses were analyzed in three ways. First the CV-structure of the recalled syllables was analyzed, counting the numbers of each type. Second, errors were categorized in terms of substitutions, deletions etc., establishing the proportions of each. Finally, the responses were analyzed in terms of the origin of their phonemes.

### 2.2.1 CV-structure

The CV-structure of each response was recorded. The results are shown in table 5.2. 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 5.2 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 Treiman and Danis (1988) data,

## 2. SIMULATIONS

---

that a small proportion of responses with other structures were produced.

### 2.2.2 Substitutions, insertions and deletions

Errors were categorized into substitutions, insertions, deletions and so on. These results are shown in table 5.3. 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 (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<sup>3</sup>.

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 5.4 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% 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

---

<sup>3</sup>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.



## 2. SIMULATIONS

---

Table 5.1: *Three examples of stimulus lists from Treiman and Danis (1988, experiment 1)*

---

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

---

Table 5.2: *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

---

lists used by Treiman (in press) contained fewer target syllables, than those used by Treiman and Danis (1988) and in the simulation<sup>4</sup>. 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.

---

<sup>4</sup>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

## 2. SIMULATIONS

---

Table 5.3: *Distribution of error types in simulations of serial recall of lists of six monosyllabic nonwords. The scheme used is similar to that of Bisiacchi et al. (1989)*

		% responses
single phoneme errors	substitutions	28.64
	deletions	3.50
	insertions	2.94
multiple phoneme errors <sup>a</sup>	substitutions	4.78
	substitution+insertion	4.61
	substitution+deletion	1.09
	deletion	0.03
	insertion	0.09
	deletion+insertion	0.09
complex errors <sup>b</sup>		0.04

<sup>a</sup>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.

<sup>b</sup>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/.

Table 5.4: *The proportion of errors with different phonemic lengths. Data from Treiman (in press) are shown for comparison*

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

### 2.2.3 Source of constituent phonemes in CVC errors

Following Treiman and Danis (1988), the model's CVC errors were categorized 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  $C_a V_b C_c$ . If the initial consonant came from one target syllable, but the remaining vowel and final consonant came from another, the response would be coded as  $C_b V_a C_a$ . If all of the phonemes in a CVC response originated from the *same* target item, it would be coded  $C_a V_a C_a$ . 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  $C_x V_a C_a$ . 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 5.5 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 5.5.

Almost all of the errors Treiman and Danis (1988) observed were CVC syllables. They placed great emphasis on the finding that most of these errors combined the onset of one item, with the rhyme of another ( $C_b V_a C_a$ ). Although this may well be an important point (see "Error mechanisms: Syllable structure and errors"), the variety of errors is perhaps even more striking – some combine the vowel of one item with

## 2. SIMULATIONS

---

Table 5.5: *Responses categorized according to the scheme adopted by Treiman and Danis, 1988 (T&D). 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 for details of error categories), 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.*

Type of response	% of all responses		
	Experiment		Simulation
	T&D(1988)	Treiman(1994)	
correct	42.64	44.17	54.18
$C_b V_a C_a$	14.68	10.59	9.13
$C_a V_a C_a$	6.16	9.58	0.48
$C_a V_a C_b$	5.32	4.27	6.94
$C_x V_a C_a$	4.74	3.96	4.67
$C_a V_b C_c$	4.43	0.80	1.42
$C_a V_b C_a$	3.49	3.33	4.62
$C_a V_x C_b$	2.87	3.75	0.54
$C_a V_x C_a$	2.15	4.83	1.58
$C_x V_a C_b$	2.04	0.80	0.89
$C_a V_b C_x$	1.85	0.69	0.30
$C_a V_a C_x$	1.53	1.15	2.27
$C_x V_x C_a$	1.10	1.18	0.14
$C_a V_x C_x$	0.51	0.90	0.07
$C_x V_a C_x$	0.48	0.24	0.21
$C_x V_x C_x$	0.26	0.38	0.02
non-CVC error	5.77	9.38	12.54

## 2. SIMULATIONS

---

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.,  $C_x C_a C_a$ ) and those involving phonemes from multiple sources (e.g.,  $C_a C_b C_c$ ,  $C_a C_b C_x$ ). 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. For instance, the model produces many more  $C_b V_a C_a$  errors than  $C_x C_x C_x$ . This is illustrated by the scatterplot in figure 5.6 a), in which for each category of response, the experimentally-observed and simulated proportions from table 5.5 are plotted against one another ( $r^2 = 0.94$  for these data sets). Figure 5.6 b) shows the same simulation data plotted, this time, against experimental data from Treiman (1994) ( $r^2 = 0.94$ ).

For completeness,  $C_a V_a C_a$  errors have been included in the analysis. Though the simulation did yield a small number of  $C_a V_a C_a$  errors, these all involved rearrangements of phonemes from the same list item<sup>5</sup>. By contrast, most, if not all of the experimentally-observed  $C_a V_a C_a$ 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  $C_a V_a C_a$  errors are considerably under-represented in the model's output (there

---

<sup>5</sup>Typically a pre-vocalic consonant is repeated in the post-vocalic position (e.g., /v3f/ → /v3v/ , /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.

---

are also correspondingly more correct responses).

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

### 2.3 Modelling 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 modelled 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 sub-normal 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 representation 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

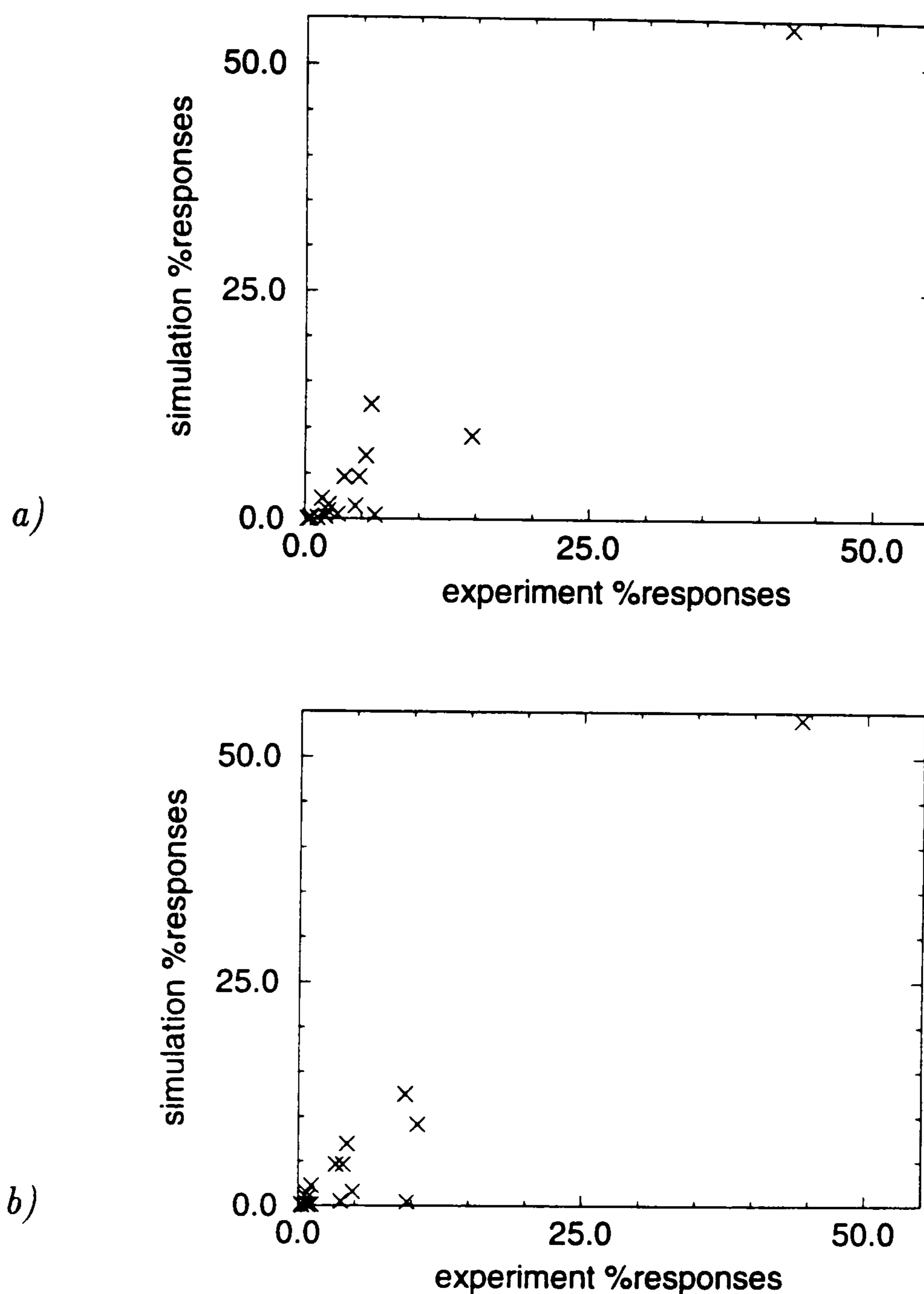


Figure 5.6: Scatterplots in which, for each response categories reported in Treiman and Danis (1988) the experimentally observed rate (averaged over subjects) is plotted against the predicted rate (from the simulation output). Experimental data are from a) Treiman and Danis (1988,  $r^2 = 0.94$ ) and 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.

## 2. SIMULATIONS

---

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 5.6 shows the distribution of errors in simulations of nonword repetition, in which the halflife of the temporary weights has been reduced from 5.0s to 3.5s. The same materials and procedure were used as in the simulations described in the



## 2. SIMULATIONS

---

section on '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 eight-year olds in the experimental studies. After it, it is close to that of 7 year old children (see table 5.7).

### 2.4 Error mechanisms

This section describes the mechanisms underlying the various types of phonemic error the model produced in simulations of Treiman and Danis' (1988) serial recall experiment: substitutions, insertions and deletions. The same mechanisms apply to the less frequent errors in repetition. These 'error mechanisms' are not put into the model simply to ensure that the model's performance is in line with existing data, they arise from the use of a syllabic structure in the short-term representation of speech. The use of structural information means that the serial order of a sequence

## 2. SIMULATIONS

Table 5.6: *The distribution of error types as a proportion of all responses, in simulations of nonword repetition. For comparison, the repetition errors made by four year old children (Gathercole et al., in press), and patients RR (Bisiacchi et al., 1989) and IGR (Caramazza et al., 1986) are shown.*

		% 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	–	–	–

Table 5.7: *Table showing the percentage of correct responses in simulations of nonword repetition for stimuli of different lengths. Mean data from children of ages 4-9 years (Gathercole, personal communication)*

Number of Syllables	% correct						
	Experiments					Simulations	
	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

## 2. SIMULATIONS

---

of phonemes in a syllable can be reduced to a simpler endogenous cyclical sequence. In addition, it ensures that output conforms to basic phonotactic constraints. This means that output will be pronounceable, and that even when errors occur, they will usually retain characteristics of the target. Because the model specifies a particular mechanism for serial output (as any implemented model must), it is possible to explore *how* errors occur.

### 2.4.1 Substitution errors

Table 5.3 shows that the majority of the errors made by the model were single phoneme substitutions, which made up 28.64% of all responses. These errors may be subdivided into two types; errors where the replacing phoneme is present in the target list (contextual substitutions), and those where it is not (non-contextual substitutions). If multiple phonemes are substituted in an error, a third category can be added to include substitutions involving both contextual and non-contextual components.

It is important to be aware that the set of contextual substitutions is likely to include some examples of errors which arise through the same mechanisms as the non-contextual substitutions: sometimes an error with a putative source in the stimulus list will arise through the same chance processes that give rise to non-contextual errors. An artifact of the categorization of errors as contextual and noncontextual is that, as more phonemes are used in the target sequence, there is correspondingly less opportunity for an error to be classified as non-contextual. This factor will clearly play a role in determining the relative frequencies of apparently contextual and non-contextual errors for different lists.

In the current model, non-contextual errors arise because an activated template node activates all of the other phonemes associated with its 'slot' in the template

## 2. SIMULATIONS

---

as well as the target phoneme. The other phonemes from the slot thus compete to replace the target phoneme. 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 of a particular phoneme, and its recall, weight decay in the content pathway will tend to reduce its advantage in the competition, so that noise in the system will begin to play an important role in determining which of the competing segments is output – occasionally, the winning phoneme will be one from outside the target list.

Contextual substitutions occur through much the same mechanism, but with an additional factor: phonemes in the target sequence also have direct excitatory connections from nodes in the syllable layer. Since the syllable nodes are activated in parallel, phonemes from the list 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 above, even when they are receiving no input from the content pathway, phonemes from within the list compete on an equal footing with segments from outside the list, so that contextual substitutions can involve phonemes from any part of the intended utterance, providing they fit the phonotactic constraints imposed by the syllabic template. 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

## 2. SIMULATIONS

---

Table 5.8: *The distribution of contextual and non-contextual substitution errors in the model's attempts to recall the first and last items in lists of 6 CVC nonwords (analysis includes only CVC errors).*

Error Type <sup>a</sup>	% responses		
	position 1	position 6	across positions
Contextual substitutions	25.47	13.40	22.59
Non-contextual substitutions	7.33	11.13	8.96
Mixed substitutions <sup>b</sup>	2.27	1.40	1.73

---

<sup>a</sup>including single phoneme substitutions and multiple phoneme substitutions, but not including substitution+deletions or substitution+insertion errors.

<sup>b</sup>multiple phoneme substitutions where one component of the substitution is contextual, and another is non-contextual.

rhyme or *vice versa* are not due to any contextual effect. Such errors are no more likely than any non-contextual substitution.

The pre-emptive activation of phonemes via the content pathway does not play any role in the last serial position of the list. Here, all the syllable units associated with prior responses have been suppressed and there are no forthcoming items to be primed. The error rates for this position show how contextual and non-contextual errors might be expected to be distributed if priming were not playing an important role. By contrast, the priming of subsequent items is always important in the first position of any list of length greater than one. The distribution of contextual and non-contextual substitutions observed in the first and last positions in lists of length six syllables are shown in table 5.8. Note that although there are still some apparently contextual errors in the last serial position, these arise through precisely the same mechanisms as the non-contextual substitutions.

## 2. SIMULATIONS

---

### 2.4.2 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, and none of the immediately forthcoming items in the list. For this reason vowel deletions are very unlikely, since the vowel slot is used in each 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:

$/\text{ʃim}/, /wɒd/ \rightarrow /ʃi/ \dots$

In order to prevent such deletions from occurring too frequently, the input to the template nodes,  $\epsilon$ , must be increased for longer lists, so that the slot nodes can still be accessed (see appendix B). As the weights decay towards zero, the slot nodes will only fire if  $\epsilon$  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 gobbledygook (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.

## 2. SIMULATIONS

---

### 2.4.3 Insertions

In the model insertions can occur in two different ways: firstly a single phoneme may be exchanged with one of the clusters (/sp/, /st/ or /sk/) which are represented by single nodes in the phoneme group, for example /sɒt/ → /skɒt/ or /paf/ → /spaf/. These errors occur in the same way as other substitutions, but are indistinguishable to the observer from insertions – they are very rare. Structural insertions may also occur, where a phoneme node's activation exceeds the competitive threshold, and is output, despite the fact that its 'slot' was not used in the target item, for example /tɪz/ → /trɪz/. Errors like these account for the vast majority of insertions.

Structural 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:

/ʃɪm/, /wɒd/ → /ʃwɪm/ , ...

A template node may also fire during the wrong syllable as a result of noise in its input. This is likely to occur as the amount of external input,  $\epsilon$ , that must be applied to the template to access it increases for longer lists. Ultimately, decay is such that a slot will only become active if  $\epsilon$  is so large that a substantial proportion of responses will include such insertions.

Insertions and deletions result largely from interactions between target syllables. These interactions produce more errors as 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 I 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.

### 2.5 Syllable structure and errors

From the preceding discussion, it should be clear that the types of errors to which a particular syllable is vulnerable are determined by three factors: 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 the basic effects of each are summarized in turn.

1. *Long-term phonological knowledge.* Because different slots in the template are associated with different numbers of phonemes, some parts of the syllable will be more vulnerable to error whenever they are accessed whatever the stimulus. Slots 1 and 5, are associated with more phonemes than are slots 2 and 4. The nodes representing phonemes associated with a particular template position are all activated whenever that position is accessed in recall. The greater the number of competing phonemes, the greater the chance that the wrong one will win the competition to be output.
2. *The target syllable's structure.* Put simply, if a particular slot is used in a target syllable, then the phoneme at that position is open to substitution or 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 (or polysyllable) influence recall by providing priming input to the syllabic template. If syllables adjacent to the target share the same structure, the priming serves to reinforce the activation of the target structure so that there is less likelihood of structural errors. However, because of the structural overlap between syllables in the list, there will be increased competition for the slots which are used, so contextual substitutions will be more frequent. Conversely, if syllables near to the target are structurally very different (use different slots), then the response will be vulnerable to structural errors, while contextual substitutions will be less frequent. Additions result from the extra input of adjacent syllables to slots not in the target syllable. Deletions come about because the activation of the slot used in the target is relatively weak, being unsupported by input from units representing neighbouring syllables.

For the stimuli used by Treiman and Danis (1988), the interaction between these factors leads to the distribution of errors shown in table 5.5. Treiman and Danis (1988) place 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  $C_aV_aC_b$  than the human subjects).

The difference between the frequencies of the two categories of error in the model's output is largely attributable to structural properties of the stimulus items (the same ones used by Treiman and Danis, 1988, experiment 1): while most of the pre-vocalic

## 2. SIMULATIONS

---

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 consonants more vulnerable to substitution.

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, this manipulation should 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. It seems likely that, while the structural properties of the stimulus items did tend to produce more substitutions involving pre-vocalic segments, this factor alone cannot account for the difference in the frequencies of  $C_bV_aC_a$  and  $C_aV_aC_b$  errors in the experimental data. Thus, a modification to the current model may be necessary. The one that suggests itself is that competition occurs at the level of the syllabic constituents (between onsets and between rhymes) rather than at the level of syllabic units (as in the current model). This account, broadly consistent with the current architecture, 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.

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 they replaced in nonword serial recall. Caramazza et al. (1986) and Bisiacchi et al. (1989) found that their subjects' substitutions tended to share a particular feature – manner of articulation. Manner of articulation is correlated with sonority. Because the sonority relations between phonemes are captured in the organization of the syllabic template in the current model, almost all of the phonemes sharing a particular manner of articulation are associated with the same slot(s) in the template. So when one phoneme substitutes for another in the same slot (point 2 above), there is a good chance that the target and intrusion share this feature value.

### 2.6 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.

Because the current model produces phonological output, which is amenable to qualitative as well as quantitative analyses, we can be a good deal more specific about the nature of the predicted effects: lists in which the syllable structure alternates

## 2. SIMULATIONS

---

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 ( $C_1V$ ,  $C_1V$ ,  $C_1V$ ,  $VC_5$ ,  $VC_5$ ,  $VC_5$ ). In alternating lists, there is less overlap between the structures of adjacent items. The following predictions can therefore be made:

1. insertions and deletions will 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 will 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.

As noted above (see table 5.8), the model predicts that the ratio of contextual to non-contextual substitutions will not be constant over the list, but that there will be a greater proportion of contextual errors at the beginning than at the end. The size of the effect is likely to be fairly small, because quite a large proportion of errors categorized as contextual substitutions are expected to arise through the same chance processes as non-contextual substitutions.

## 6. Future directions and conclusions

---

### 1 Extending the model: short-term memory and the lexicon

The performance of the *sonority wheel model* described in chapter 5 shows many of the characteristics of human subjects engaged in short-term memory tasks involving nonwords. The model does this without any contribution of long-term lexical-phonological knowledge. However, studies showing effects of wordlikeness on nonword repetition (Gathercole et al., 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 nonwords, the sonority wheel model postulates phonological retrieval processes common to speech production and short-term memory. This development makes it possible to extend the

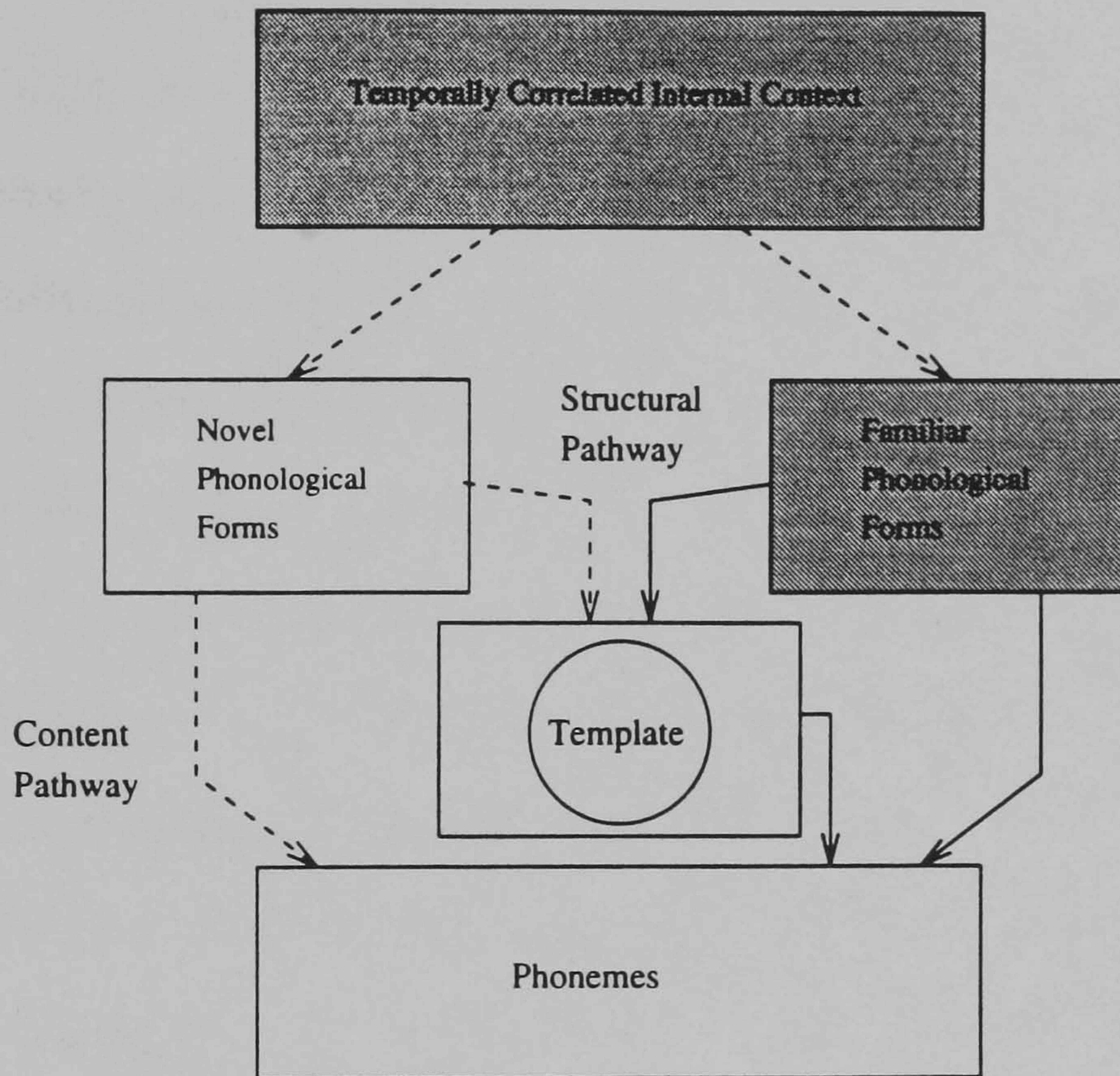


Figure 6.1: *Extending the existing architecture to include the representation of long-term lexical-phonological knowledge. The shaded areas represent structures not implemented in the version of the model described in chapter 5. Note that dashed lines represent pathways made up of temporary weights, the solid lines represent connections which do not decay over time.*

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 figure 6.1 (cf. figure 5.1). The phonological information required to produce familiar words is represented in the same way as that used in the simulations reported in the previous chapter, except that the connections which hold the information are *permanent* rather than temporary. 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. Figure 6.2 shows how a familiar monosyllable (“barn”) would be represented. Like Dell (Dell, 1986), we must assume that in speech production as in short-term memory, forthcoming words are activated in advance of their articulation to some degree. As in the recall of nonwords, phonemes from the same slot are activated simultaneously during retrieval. In this mode of operation, the model will function in a similar manner to Dell’s model of sentence production (Dell, 1986). Thus when articulating the phrase barn door, the /b/ and /d/ will compete with one another. As Dell demonstrated, such competition can provide a compelling account of the pattern of errors in speech production.

It is further suggested that the permanent weights are established by means of an iterative procedure which exploits the temporary pathways. Once a novel word has been recognised as such, this could be achieved by rehearsing the material held in the temporary weights in short-term store, while activating an unassigned syllable unit in the long-term phonological store. Hebbian learning could be used to gradually strengthen connections between the syllable unit and nodes representing its

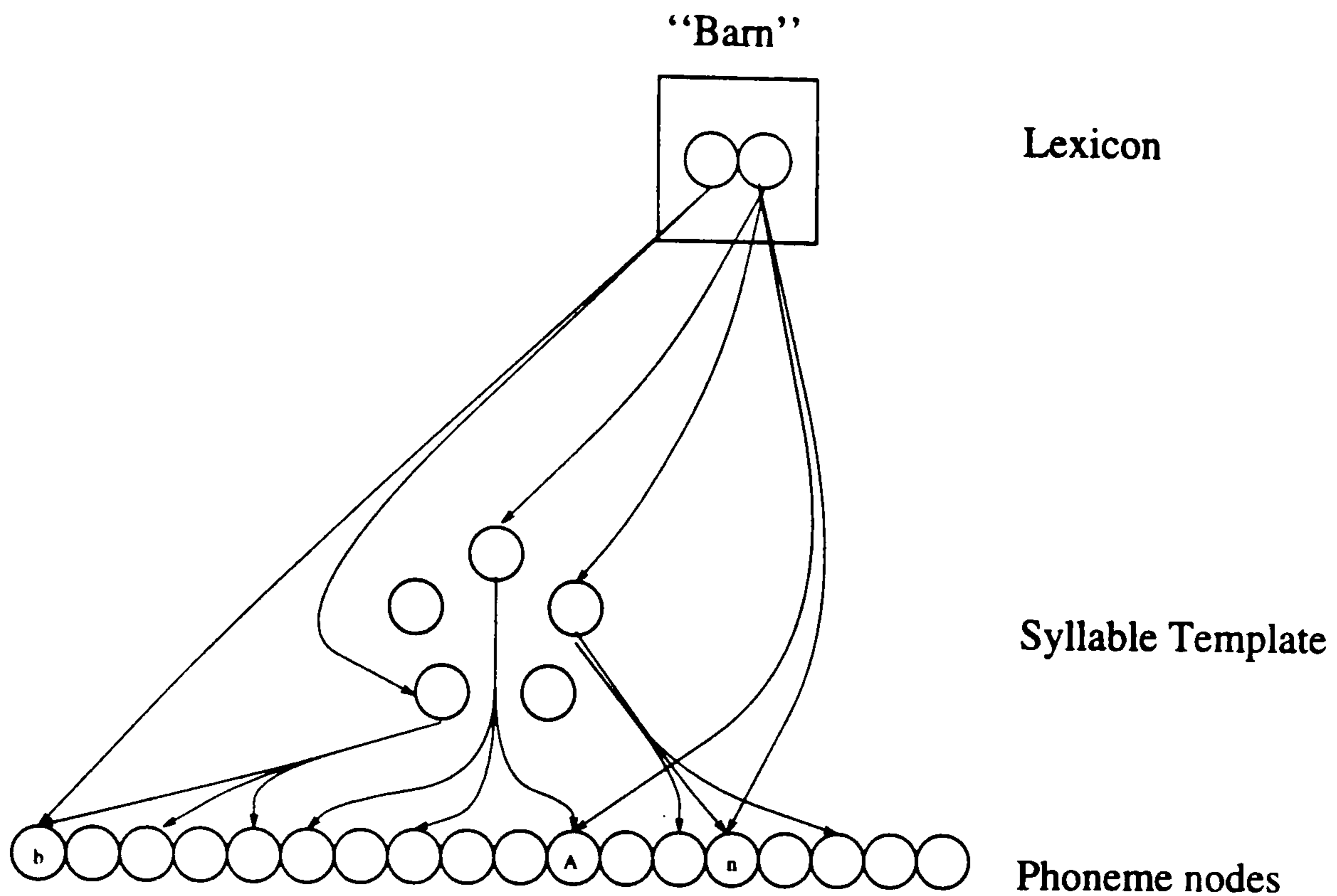


Figure 6.2: *This figure (cf. figure 5.2) shows in detail how a familiar monosyllable (“barn”) would be represented in the phonological lexicon. Note that all the connections are permanent.*

constituent phonemes and ‘slots’ (the same kind of mechanism is used by Burgess & Hitch, 1992 to refresh the temporary weights in their model). Thus the acquisition of new vocabulary will depend on the efficiency of the system responsible for short-term storage of phonological forms in line with recent experimental data (Gathercole & Baddeley, 1989).

A global increase in the rate of decay in the temporary weights of the network will cause more phonological misorderings for nonwords, as is the case in the simulation described in chapter 5 section 2.3. Performance on word recall, however, will not be similarly affected, because the connections in this pathway are permanent. In contrast to nonwords, the representation of the detailed phonological form of words is not subject to decay. However, item order information is: the connections between the internal context units and syllable units representing both words and nonwords are temporary ones (as in Burgess & Hitch, 1992). This means that a global increase in decay results in poorer memory for the serial order of familiar



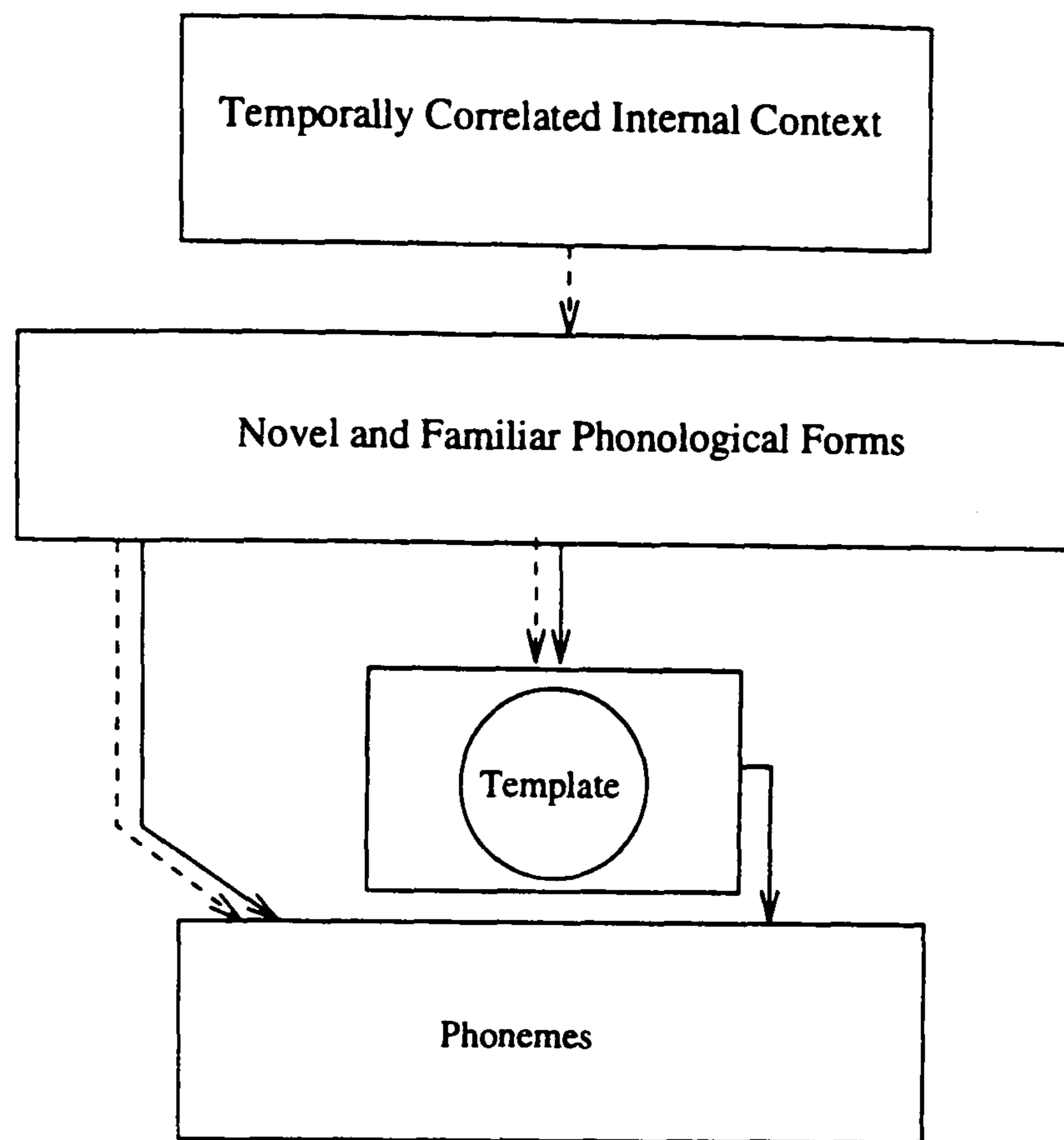


Figure 6.3: *An alternative architecture does not explicitly differentiate between nodes used to represent familiar and unfamiliar phonological forms. Instead, weights are subject to dual potentiation, having both permanent and temporary components (the adjacent dotted and solid lines arrows represent such dual weights). With repeated use, the permanent component gradually builds up. However it is still necessary to be able to distinguish unassigned nodes (with no permanent lexical association) so that they can be recruited to become associated with new words when they are first encountered.*

items (e.g., subnormal digit span) while their phonological forms are preserved. The same manipulation results in an impairment of memory for novel phonological forms (e.g., poor nonword repetition). Thus it is suggested that a single global manipulation (variation in decay parameter) differentially affects the model's performance on words and nonwords. This is a more parsimonious explanation of the nature and effects of damage than current accounts which assume that observed dissociations reflect the lesioning of independent word- and nonword- output systems (e.g., Caramazza et al., 1986).

Another possibility, raised by Burgess (Burgess, 1995), is that the same units are used to represent both words and nonwords, with weights having both permanent and temporary components. The temporary component of the weight is rapidly strengthened but decays quickly, while the permanent component gradually builds up with repeated exposure to the stimulus, or through rehearsal. This proposal is compatible with the current framework (see figure 6.3, the only modification being that there is no distinction between syllable units used to encode words and nonwords. It does require an additional mechanism for choosing an unassigned node when a new stimulus is encountered.

Wordlikeness effects on nonword recall can also be understood within either framework if it is assumed that when learning unfamiliar words, nodes in the familiar-syllable group are partially activated by wordlike items. So for instance, presentation of the nonword /dʒɒp/ would partially activate syllable nodes corresponding to the familiar syllables “shop” and “top” etc. These syllable nodes become associated with a particular context state, and are thus 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 additional input to the template and phoneme groups, where it augments the input from the temporary pathways, lessening the impact of decay on the short-term representation.

## 2 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 in chapter 2, the model is built 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 in chapter 3 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.

### 2.1 Models of serial order and short-term memory

Many models in the fields of short-term memory and language have little to say about the nature of serial learning and recall, preferring instead to rely on the use of vague terms like 'buffer' to imply the existence of mechanisms capable of serial processing without specifying them in any detail. Such models can thus offer no satisfactory account of the data addressed in chapter 5, a fact which should serve to indicate the need for a lower-level approach to modelling in areas where serial processing is involved.

Some theories suffer from a reliance on associative chaining which cannot deal comfortably with sequences containing repeated items, and does not predict the prevalent error types. In particular, associative chaining models cannot account for anticipatory errors and exchanges.

One of the most important theoretical constraints on a short-term memory model is that it must be capable of single-trial sequence learning. Some models of serial order can be ruled out because they do not satisfy this constraint. For instance, models employing variants of the backpropagation learning algorithm (e.g., Jordan, 1986; Elman, 1990) 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 behaviour in short-term memory tasks.

## 2. THE MODEL IN RELATION TO OTHER WORK

---

P-I models of serial order do not fall prey to the same criticisms. The most promising general theory in the area is Competitive Queuing. CQ models use a time-varying control signal to represent serial position. By forming associations between successive items in a sequence, and successive states of the dynamic signal, a representation of a sequence can be built up in a single presentation. Repeated items can be dealt with in principle, because each item can be associated with more than one state of the control signal. During recall, the signal is reconstituted, and this activates the items in series. The performance of CQ models is critically dependent upon the nature of the control signal. If the signal is rich enough, there may be a unique, orthogonal state of the control signal for every item to be stored – in this extreme case, the items would be activated one at a time during recall, and performance would be perfect. There would be little to choose between such an account and a vague reference to a buffer. Error data could certainly not be explained. However, it is generally assumed that there are limits on the dimensionality of the control signal, and it follows that there will necessarily be some similarity between states of the control signal at different times. This leads to the potential for errors to occur as items activated in parallel compete to be output. Indeed a control-signal of very low dimensionality may not be able to store sequences much longer than a few items.

It is generally assumed that the control signal changes gradually over time, that is that adjacent states of the signal are more similar to one another than are well-separated states. There is thus a tendency for items to be active in parallel if they occurred close to one another in the learned sequence. This parallel activation of adjacent items provides a natural account of anticipations and exchange errors. However, the success of CQ models in accounting for the pattern of common error types has, to date, been confined to tasks where the serial order of the items to be learned is not constrained. For example, in a memory span task, the order of the items is arbitrary, and a CQ model of the articulatory loop predicts that errors will largely involve the local transposition of items, often in exchanges. This prediction

is to be borne out by empirical data (Henson et al., in press).

In the recall of novel phonological forms, there are strong constraints on the ordering of phonemes. Thus, the ordering mechanisms which apply to unconstrained sequences cannot simply be extended to the phonological level as Glasspool (1994) suggests (see discussion in chapter 4). Although Glasspool's model is capable of dealing with both words and nonwords, and predicts some of the effects described by Hulme et al. (1991), it is not capable of dealing with the fine-grain error data from nonword recall addressed in this thesis. The same criticisms could be applied to any linguistically-unconstrained model which might be put forward; 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 sonority wheel model accounts for the phonological regularities in recall errors, and is in other respects consistent with Burgess and Hitch's (1992) 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.

### 2.2 Models of speech production

In comparing the sonority wheel 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 et al., 1993) have generally used backpropagation learning or related algorithms which require multiple exposures to a sequence in order to learn it. In the context of short-term memory, however, such iterative learning procedures are not an option: any information which is retained about an syllable's phonemic content and serial structure must be captured in a single exposure, as the stimulus occurs. Learning phonemic content in a single trial is not problematic, it simply requires that successive phonemes become associated with a single node standing for the target syllable. Learning the serial structure so rapidly is more difficult. In the current model, this is achieved by associating the structure of each target syllable with a simple reproducible sequence, which represents well-learned constraints on syllable structure. In contrast to Dell et al. (1993), this approach 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, the sonority wheel 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 the phonological contents of that frame.

### 2.3 Syllable structure

The sonority wheel 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 proposes 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 modelling 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; I 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<sup>1</sup>. Speech production models can be consistent with syllabi-

---

<sup>1</sup>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 et al., 1994).

## 2. THE MODEL IN RELATION TO OTHER WORK

---

fication theory by simply specifying frames which have the desired properties. They are free to do so because, in the main, such models do not attempt 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 identity 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.

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 specified in chapter 4. However, 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.

Whether or not one regards the specific template proposed in the current model as realistic, the basic 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. However, within the framework of syllable-schema models of verbal short-term memory, a number of possible templates are still possible. One might for example suggest a template with more or fewer ‘slots’, or a distributed pattern of activation rather than local representation of syllable position/sonority. The phonemic representation of input and output could be replaced by one using larger or smaller units (onset/rhyme clusters, features). In this area, theoretical work would



benefit from further experimental input. 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. I hope that the model put forward here will serve to stimulate such research.

## 3 Conclusions

In this thesis I have set out to provide a formal model of one of our most basic linguistic behaviours – the recall of unfamiliar words. Analysis of the empirical literature showed patterns of error which could not be explained by existing models of serial memory. The ordering mechanisms of these models are quite general, and make little or no reference to the special constraining principles which apply to verbal stimuli. As a result they predict that ordering errors will have the same character whether the units being sequenced are letters or digits in an arbitrary sequence, words in a sentence, phonemes in a word, or whatever. However, although the sequence of digits in a span task is arbitrary, the sequence of speech sounds in a nonword is not. Nonwords are constrained to fit basic phonological constraints (such as the sonority principle) and so, it appears, are the errors subjects make in recalling them. The serial order of the speech sounds is represented in terms of the linguistic structure of the utterance. It was therefore necessary to develop a model which could capture the important aspects of the linguistic structure of an utterance, as it was being perceived. This necessity led to the development of the syllable template described in chapter 4. Using the syllable template, it was possible to extend Burgess and Hitch's model of the articulatory loop to deal with unfamiliar sequences of phonemes in a plausible manner. Simulations showed the characteristic patterns of error which are observed experimentally. This pattern of

### 3. CONCLUSIONS

---

errors reflects a system of phonological constraints which are consistent with those applying to spontaneous speech errors.

The bringing together of theoretical insights from models of speech production, serial order and short-term memory results in a fairly complex model, but one with a much greater scope than its precursors. The sequencing mechanism suggested in the new model for nonwords could equally be used for the phonological retrieval of words. The new model, unlike models of spontaneous speech production, shows how novel words can be learned in a single presentation. Speech production and verbal memory (short- or long-term) can profitably be regarded as different functions of the same integrated system from the learning, representation and recall of speech.

## **A. Coding used in phonological simulations**

---

Table A.1: A table showing the form of local coding used for phonological information in the simulations reported in chapter 5. Also shown are the template slots that are activated (during learning) by each 'phoneme'.

## 1 Vowels

phonetic symbol	example	phoneme node	associated template slot(s)
i	key	0	3
a	car	1	3
ɔ	core	2	3
u	coo	3	3
ɜ	cur	4	3
ɪ	pit	5	3
e	pet	6	3
æ	pat	7	3
ʌ	putt	8	3
ɒ	pot	9	3
ʊ	put	10	3
ə	about	11	3
eɪ	bay	12	3
aɪ	buy	13	3
ɔɪ	boy	14	3
əʊ	go	15	3
aʊ	cow	16	3
ɪə	peer	17	3
eə	pear	18	3
ʊə	poor	19	3

---

## 2 Consonants

phonetic symbol	example	phoneme node	associated template slot(s)
w	wet	20	2
j	yet	21	2
l	led	22	2, 4
r	red	23	2, 4
m	map	24	2, 4
n	nap	25	2, 4
ŋ	hang	26	4
p	pea	27	1, 5
t	tea	28	1, 5
k	cap	29	1, 5
f	fat	30	1, 5
θ	thing	31	1, 5
s	sip	32	1, 5
ʃ	ship	33	1, 5
h	hat	34	1
b	bee	35	1, 5
d	doe	36	1, 5
g	gap	37	1, 5
v	vat	38	1, 5
ð	this	39	1, 5
z	zip	40	1, 5
ʒ	measure	41	1, 5
dʒ	gin	42	1, 5
tʃ	chin	43	1, 5
sp	spit	44	1, 5
st	rest	45	1, 5
sk	disc	46	1, 5

# B. Technical appendix to chapter 5

---

This appendix specifies the mathematical details of the model described in chapter 5 and gives the parameter values used in the simulations reported there.

## 1 Activation

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

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

The term  $\nu$  is a random variable (representing noise) drawn from a Gaussian distribution of mean 0.0, and standard deviation  $\sigma$  (set to 0.065 in the reported simulations). The parameter  $\tau$  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' behaviour is implemented by setting  $\tau$  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

$$I_i = \sum_j \max(0, a_j) w_{ji} \quad (2)$$

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.

## 2 Learning

The strengths of the temporary connections are developed by exposure to a stimulus sequence according to equation 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.

## 3 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 are 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 3, where  $s_i$  designates the syllable node associated with the  $i^{th}$  target syllable, and the current target is the  $j^{th}$ :

$$a_{s_i} = \begin{cases} f(I_{max}(1 - \frac{i-j}{r})) & \text{in the range } j \leq i < j + r \\ 0 & \text{outside it.} \end{cases} \quad (3)$$

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  $\epsilon$  cyclically to each node in the template group in turn. Thus, the net input to a template node (designated  $u_k$ ) becomes:



### 3. RECALL

---

$$I_k = \sum_i a_{s_i} w_{ik} + \epsilon \quad (4)$$

The value of  $\epsilon$  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  $\epsilon$  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  $\epsilon$  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  $\epsilon$  is too great. In order to prevent this occurring, the value of  $\epsilon$  must be such that with input from the primed syllable nodes ( $s_{j+1}$  thru  $s_{j+r-1}$ )  $I_k$  is less than 1.0. This gives an upper limit to the value of  $\epsilon$ :

$$\epsilon_{max} = 1.0 - \sum_{i=j+1}^{j+r-1} a_{s_i} w_{s_i k} \quad (5)$$

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  $\epsilon$ , which must be sufficient to fire a template node  $u_k$  even if its input comes solely from the target syllable  $s_j$ :

### 3. RECALL

---

$$\epsilon_{min} = 1.0 - a_{s_j} w_{s_j,t} \quad (6)$$

If it is assumed that both insertions and deletions are equally unfavourable, the optimal value of  $\epsilon$  is half way between these limits. For a given set of parameters, the individual terms in equations 5 and 6 can be calculated for a notional template node connected to the appropriate notional syllable units, and an approximation of the optimal value of  $\epsilon$  determined: weights are assumed to have an initial strength of  $\alpha$ . 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:

$$\delta_{i-j} = 0.5^{\frac{l-(i-j)p}{h}} \quad (7)$$

where  $l$  is the list duration, and  $p$  is the duration of pauses between syllables and  $i \geq j$ . For  $r = 3$  and  $\alpha = 1$ , the optimal value of  $\epsilon$  is approximated by:

$$\epsilon_{opt} \approx \frac{2 - f(I_{max})\delta_0 - f(\frac{2I_{max}}{3})\delta_1 - f(\frac{I_{max}}{3})\delta_2}{2} \quad (8)$$

Since the weights from syllable units to template nodes decay over time, the optimal value of  $\epsilon$  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 tends towards zero and both upper and lower limits

on  $\epsilon$  converge on 1.0. As  $\epsilon$  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 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 node 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.

## 4 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, parts 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

#### 4. RATE OF PRESENTATION AND RECALL

---

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 seconds. In modelling 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 seconds was allowed between presented syllables during which weight decay continued. In simulations of nonword repetition, there were no pauses during learning or recall.

## C. Exception words

---

The syllabic template (chapter 4) parses 97% of words in a sizable corpus into the correct number of syllables, although the precise location of the syllable boundaries differs for around half. There is, however, a small subset of words which cannot be parsed into the correct number of syllables, and these are listed in this appendix. Standard syllabifications are taken from the Oxford Psycholinguistic Database (Quinlan, 1992).

EXCEPTION WORDS

Word	Phonetic Spelling	Standard	Template
ACCEPT	əksept	ək—sept	ək—sep—t
ACT	ækt	ækt	æk—t
ADOPT	ədɒpt	ə—dɒpt	əd—ɒp—t
AFFECT	əfekt	ə—fekt	əf—ek—t
APT	æpt	æpt	æp—t
ARCHITECT	ɑkitekt	ɑ—ki—tekt	ɑk—It—ek—t
ATTEMPT	əttempt	ə—tempt	ət—emp—t
ATTRACT	ətɹækt	ə—trækt	ət—ræk—t
BOX	bɒks	bɒks	bɒk—s
COLLAPSE	kəlæps	kə—læps	kəl—æp—s
COLLECT	kəlekt	kə—lekt	kəl—ek—t
COLLECT	kɒlekt	kɒ—lekt	kɒl—ek—t
COMMISSIONER	kəmiʃNə	kə—mi—ʃN—ə	kəm—Iʃ—nə*
CONDUCT	kəndʌkt	kən—dʌkt	kənd—ʌk—t
CONDUCT	kɒndækt	kɒn—dækt	kɒnd—æk—t
CONNECT	kənekt	kə—nekt	kən—ek—t
CONTACT	kəntækt	kən—tækt	kənt—æk—t
CONTACT	kɒntækt	kɒn—tækt	kɒnt—æk—t
CONTRACT	kəntɹækt	kən—trækt	kənt—ræk—t
CONTRACT	kɒntɹækt	kɒn—trækt	kɒnt—ræk—t
CORRECT	kərekt	kə—rekt	kər—ek—t
CRAFT	kraft	kraft	kraf—t
CREPT	krept	krept	krep—t
DEPTH	depθ	depθ	dep—θ
DIRECT	dɪrekt	dɪ—rekt	dɪr—ek—t
DISTRICT	dɪstrɪkt	dɪs—trɪkt	dɪst—rɪk—t
DRIFT	dɹɪft	dɹɪft	dɹɪf—t
ECONOMICS	ɪkənɒmɪks	ɪ—kə—nɒ—mɪks	ɪk—ən—ɒm—ɪk—s
EFFECT	ɪfekt	ɪ—fekt	ɪf—ek—t
ELECT	ɪlekt	ɪ—lekt	ɪl—ek—t
ERECT	ɪrekt	ɪ—rekt	ɪr—ek—t
EXACT	ɪgzækt	ɪg—zækt	ɪg—zæk—t
EXCEPT	ɪksept	ɪk—sept	ɪk—sep—t
EXCHANGE	ɪkstʃeɪndʒ	ɪks—tʃeɪndʒ	ɪk—stʃ—eɪndʒ
EXPECT	ɪkspekt	ɪks—pekt	ɪk—spek—t
FACT	fækt	fækt	fæk—t
FIFTH	fɪfθ	fɪfθ	fɪf—θ
FILM	fɪlm	fɪlm	fɪl—m
FIX	fɪks	fɪks	fɪk—s
FOX	fɒks	fɒks	fɒk—s
GIFT	ɡɪft	ɡɪft	ɡɪf—t
INSTINCT	ɪnstɪŋkt	ɪn—stɪŋkt	ɪnst—ɪŋk—t
INTERRUPT	ɪntərəpt	ɪn—tə—rʌpt	ɪnt—ərə—rʌp—t
IT'S	ɪts	ɪts	ɪt—s
ITS	ɪts	ɪts	ɪt—s
KEPT	kept	kept	kep—t
KNEW	nju	nju	n—ju

EXCEPTION WORDS

Word	Phonetic Spelling	Standard	Model
LAYER	leɪə	leɪə	leɪ—ə
LEFT	left	left	lef—t
LIFT	lɪft	lɪft	lɪf—t
MIDST	mɪdst	mɪdst	mɪd—st
MIX	mɪks	mɪks	mɪk—s
MULE	mjuːl	mjuːl	m—juːl
MUSIC	mjuːzɪk	mju—zɪk	m—juːz—ɪk
MUSICAL	mjuːzɪkəl	mju—zɪ—kəl	m—juːz—ɪk—əl
NEGLECT	nɪglect	nɪ—glect	nɪg—lek—t
NEW	nju	nju	n—ju
NEWS	njuːz	njuːz	n—juːz
NEWSPAPER	njuːspɛɪpə	njuː—pɛɪ—pə	n—juːsp—ɛɪp—ə
NEXT	nekst	nekst	nek—st
NUMEROUS	njuːmərəs	nju—mə—rəs	n—juːm—ər—əs
OBJECT	əbdʒekt	əb—dʒekt	əb—dʒek—t
OBJECT	ɒbdʒɪkt	ɒb—dʒɪkt	ɒb—dʒɪk—t
OURSELVES	aʊəsɛlvz	aʊ—ə—sɛlvz	aʊ—əs—ɛlv—z
PERFECT	pəfekt	pə—fekt	pəf—ek—t
PERFECT	pɜːfɪkt	pɜː—fɪkt	pɜːf—ɪk—t
PERHAPS	pəhæps	pə—hæps	pə—hæp—s
POLITICS	pəlɪtɪks	pɒ—lɪ—tɪks	pɒl—ɪt—ɪk—s
PRISONER	prɪzənə	prɪ—zɪ—nə	prɪz—nə*
PRODUCT	prɒdʌkt	prɒ—dʌkt	prɒd—ʌk—t
PROJECT	prədʒekt	prə—dʒekt	prədʒ—ek—t
PROJECT	prɒdʒekt	prɒ—dʒekt	prɒdʒ—ek—t
PROSPECT	prəspekt	prəs—pekt	prəsp—ek—t
PROSPECT	prɒspekt	prɒs—pekt	prɒsp—ek—t
PROTECT	prətɛkt	prə—tɛkt	prət—ek—t
REFLECT	rɪflekt	rɪ—flekt	rɪf—lek—t
RELAX	rɪlæks	rɪ—læks	rɪl—æk—s
RESPECT	rɪspekt	rɪs—pekt	rɪsp—ek—t
SELECT	sɪlekt	sɪ—lekt	sɪl—ek—t
SEX	seks	seks	sek—s
SHIFT	ʃɪft	ʃɪft	ʃɪf—t
SIFT	sɪft	sɪft	sɪf—t
SIX	sɪks	sɪks	sɪk—s
SLEPT	slept	slept	slep—t
SOFT	sɒft	sɒft	sɒf—t
SUBJECT	səbdʒekt	səb—dʒekt	səb—dʒek—t
SUBJECT	sʌbdʒɪkt	sʌb—dʒɪkt	sʌb—dʒɪk—t
SUSPECT	səspekt	səs—pekt	səsp—ek—t
SUSPECT	sʌspekt	sʌs—pekt	sʌsp—ek—t
SWEPT	swept	swept	swep—t
SWIFT	swɪft	swɪft	swɪf—t
TAX	tæks	tæks	tæk—s
THEMSELVES	ðəmsɛlvz	ðəm—sɛlvz	ðəms—ɛlv—z

## EXCEPTION WORDS

---

Words marked with an asterisk were encoded in the database with syllabic consonants (L or N) which the model does not use. Most of these can be handled simply by recoding the consonant as l or n. However, for the marked words this results in a reduction in the number of syllables recognised by the parser.



# D. Dynamics of the syllabic template

---

## 1 Dynamics of the syllabic template

The syllable template described in chapter 4 has some properties which will prove very useful in modelling short-term memory for nonwords – it captures the basic internal structure of each syllable, and is capable of parsing continuous speech. But as outlined above, its specification is not very consistent with the demands of the connectionist modelling framework which, I have argued, is the most appropriate one for addressing theoretical issues around serial order. The main problem is that the “wheel” has an inherently cyclical character, and as yet, I have provided no account of how such dynamics might arise in a simple neural network. The remainder of this chapter explores two alternative accounts of how serial activation of nodes in a cyclical structure like the one I have described could in principle be yoked to changes in the input stream. In the first account, I specify a recurrent neural network which

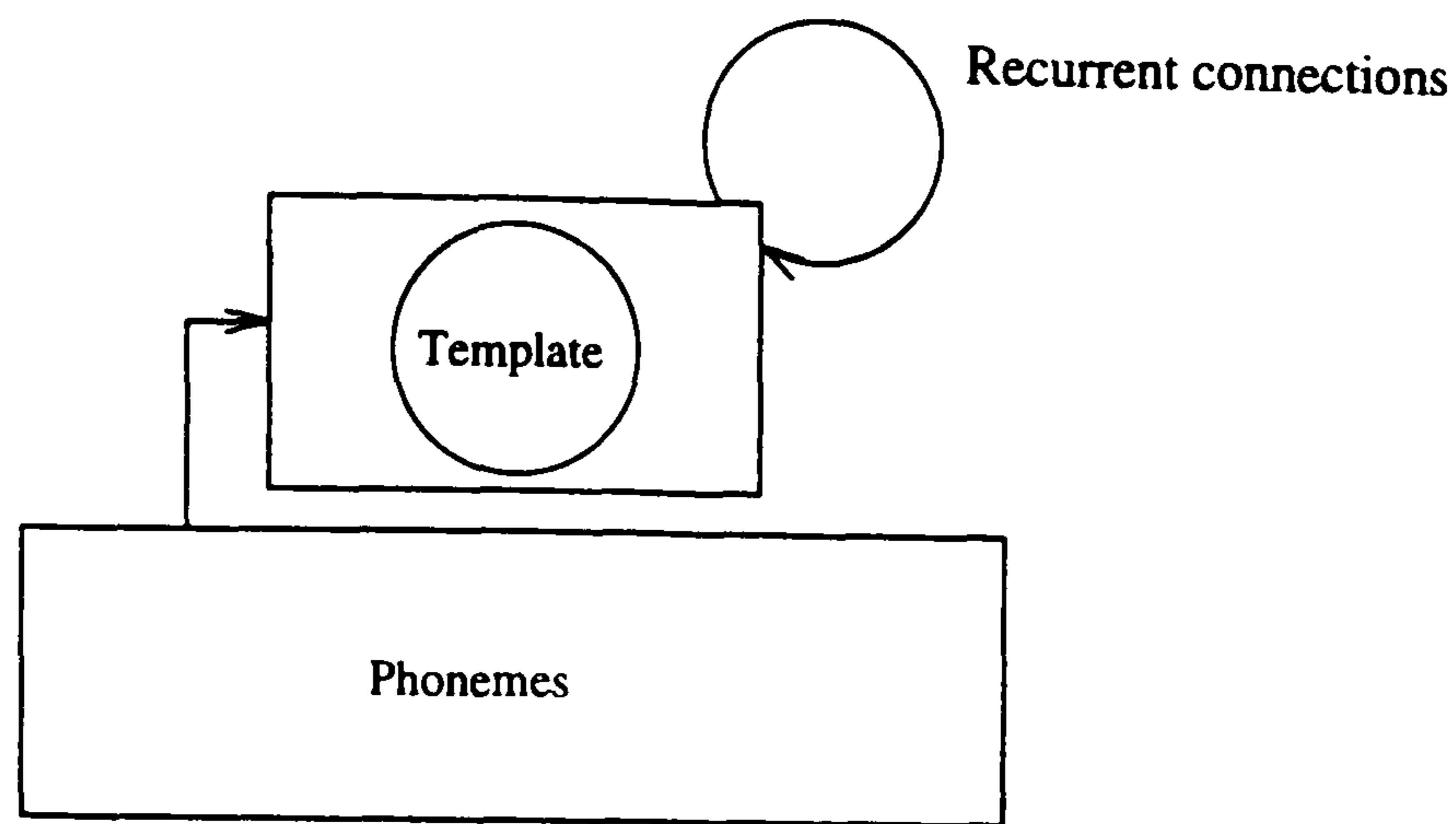


Figure D.1: *Architecture for a neural network capable of parsing syllables on-line according to the scheme developed in chapter 4*

behaves very much as the abstract formulation of the template above would suggest, acting on an input stream composed of phonological segments. In the second, I indicate how, at an algorithmic level, a measure of *syllabic phase* can be derived directly from very low-level intensity information in speech.

### 1.1 Capturing syllable structure in a recurrent network

In section 4, I showed how a simple cyclical template could be used to represent most legal syllable structures, and to parse continuous speech into syllable sized chunks. The cyclical nature of the template is of course dependent implicitly on the operation of a set of rules specifying the state of the template in every possible context. Since some segments are associated with more than one slot, the state of the template is dependent upon not only on the current state of the input, but also on the prior state of the template. This necessitates the use of a recurrent architecture like that indicated in figure D.1.

The network is composed of two groups (or layers) of nodes, the phoneme group, and the template group. In the phoneme group each of the 47 nodes represents a

## 1. DYNAMICS OF THE SYLLABIC TEMPLATE

---

particular phonological segment (see appendix A), and is activated whenever that segment is perceived. There are five nodes in the template group standing for the five 'slots' in the syllable template. All of the phoneme nodes are connected to all of the template nodes. In addition, recurrent connections in the template group link each node to itself, and to all of the other template nodes. All of the weights in the network are learned, and at the outset have small random values.

Recurrent networks like this can be trained using a version of the backpropagation 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<sup>1</sup>. 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 of less than eight 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 described in chapter 5. Briefly, each input pattern was a vector with one non-zero element standing for the currently presented phoneme. Training vectors were

---

<sup>1</sup>In this case only three previous timesteps were used.

---

Table D.1: *Table showing activations of template nodes during presentation of an unfamiliar phonological sequence*

phoneme	template 'slot'				
	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

prepared using the principle summarised in figure 4.3 in chapter 4, so that for each phoneme in the input stream the template node standing for the next available clockwise slot matching that phoneme 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 in section 4. As each phoneme appeared, the next available clockwise matching 'slot' in the template group became very active, while the other nodes remained inactive. Table D.1 shows the activations of the five nodes during the presentation of the (unlearned) sequence /slædɪŋ/. As can be seen the mechanism parses the novel input into the syllables

/slæd/ and /ɪŋ/; for each segment, a single template node (representing the next available clockwise slot in figure 4.3) is highly activated, all of the other nodes have activation values close to zero.

The connections which produce this behaviour are represented schematically in figure D.2. White boxes represent inhibitory weights, black boxes 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

## 1. DYNAMICS OF THE SYLLABIC TEMPLATE

---

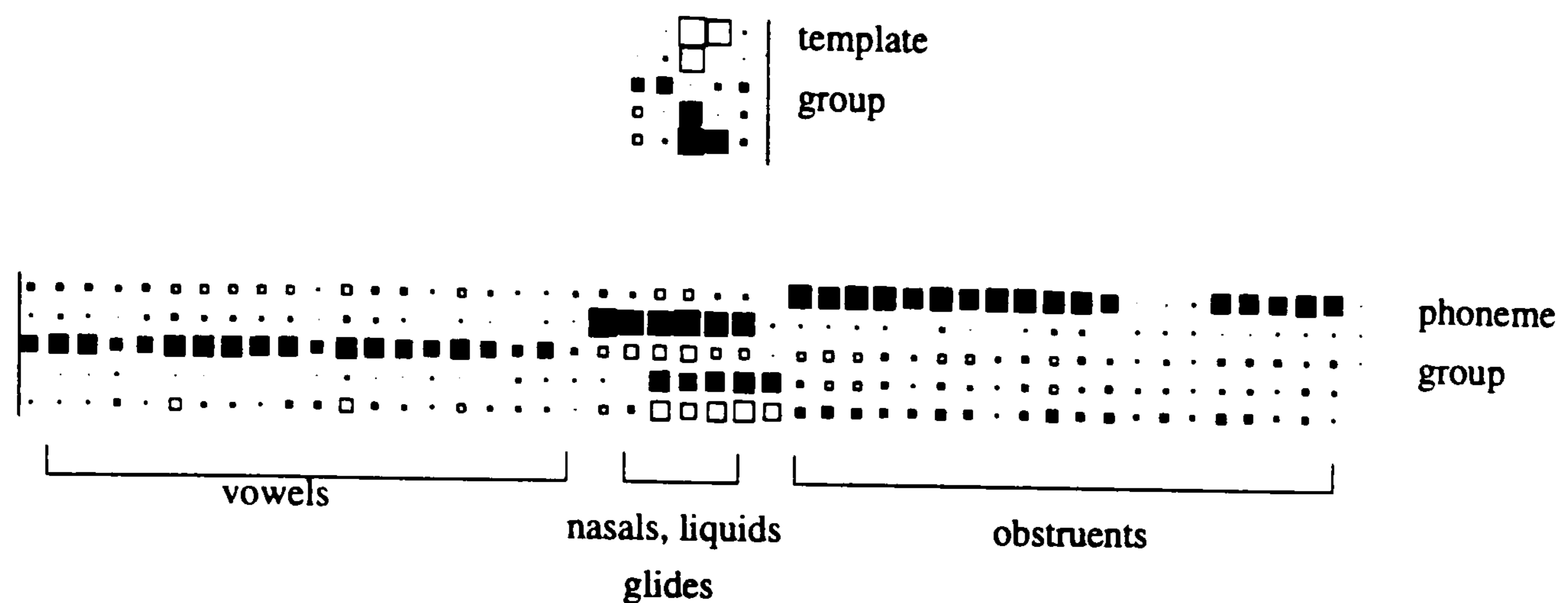


Figure D.2: . Diagram showing weights developed by the recurrent network following repeated exposure to a set of English words. The size of the square represents the strength of the connection. Black squares represent excitatory connections, white squares inhibitory ones.

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 one 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 tendency 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

## 1. DYNAMICS OF THE SYLLABIC TEMPLATE

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 position.

The recurrent network described here, behaves very much like the cyclical template described in more abstract terms in section 4, changing state cyclically in response to changes in a phonetic input stream. It is important too, that the behaviour generalizes to unfamiliar phonological sequences, i.e., sequences that were not presented during training. However, the process of training itself, relying as it does on an explicit representation of the desired output, is not very psychologically plausible developmentally. Children hear many spoken syllables, but are not provided with specific instructions as to how they should be analysed.

One response to this problem is to suggest that the process of training the network is a means to an end, the discovery of a network which is capable of behaving like the postulated template. The question of how such constraints are acquired can be investigated later. Nonetheless, my account of the dynamics of a syllable template would be more convincing, if it were possible to describe an algorithm capable in principle of specifying the template dynamics from low-level information in the speech signal itself. If cyclical syllable structure can be determined from acoustic information, it could be used to provide a teaching input to the network described above, or alternatively to specify the dynamics of the template directly from the acoustic environment. To this end in the next section, I explore a speech-based computational approach to the analysis of syllable structure which yields just such a specification. This will clearly be a departure from the more abstract approach to speech (as a phonemic or phonetically-coded stream) taken in the rest

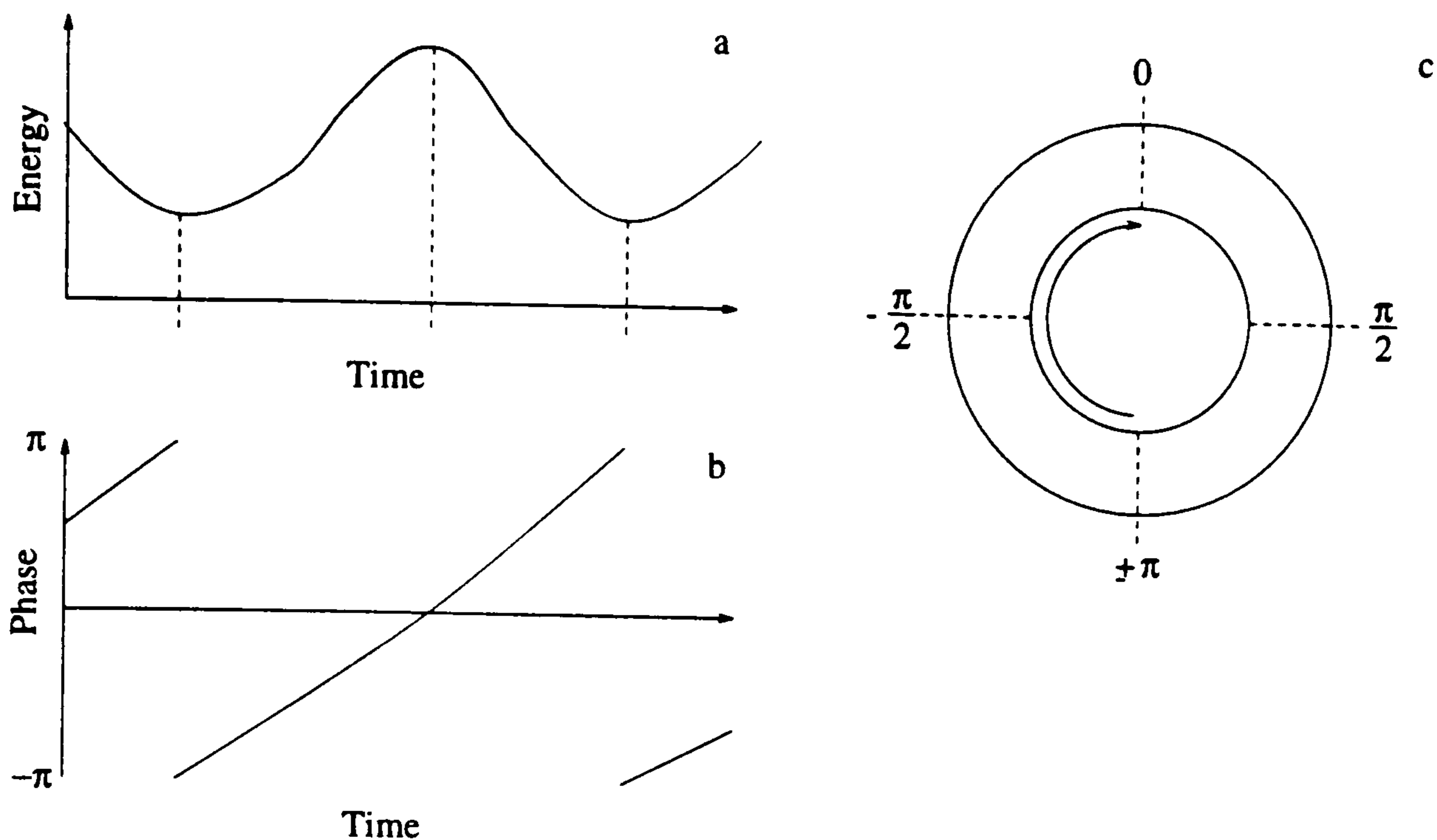


Figure D.3: *The concept of phase as applied to a hypothetical syllabic cycle.*

of the thesis.

## 1.2 Determining syllabic phase

We can regard any cyclical process, such as the syllabic-parsing process discussed above, as having a *phase* which specifies its current state with reference to an arbitrary point in the cycle. The network described above, in activating a particular series of discrete slots in a syllabic template in response to a sequence of phonological segments, determines a discrete approximation, at each point in time, to the phase of a 'syllabic cycle' – or, to coin a phrase, the *syllabic phase* of the input. The following definition may be useful: syllabic phase is the state of a hypothetical cycle associated with the perception (and production – see chapter 5) of speech (and possibly with other rhythmic sounds). In this section, I describe an analogue approach to the problem of determining syllabic phase using only low-level intensity information.

The nature of the association between slots and segments in both the abstract

formulation of the template, and the recurrent network described above, is largely a reflection of the operation of sonority constraints on phonological sequences. Because sonority is closely related to the energy or intensity of the segments concerned, it may in principle be possible to determine the desired state of a cyclical template directly from low-level information about the local time course of acoustic intensity in the signal.

The central idea behind the proposal is summarised in figure 1.2 (cf. figure 4.1). The proposed algorithm is based on a number of simplifying assumptions about the envelope of the speech signal. To the extent that these assumptions hold true for a particular utterance, the algorithm will provide useful information about syllabic phase. The simplifying assumptions are as follows:

1. That the single peak in sonority associated, as a general rule, with each syllable, is correlated with a peak in the energy of the sound.
2. That such peaks occur more or less regularly (i.e., the rate of production does not vary much) in continuous speech. The regular changes in energy at a particular characteristic frequency, (the reciprocal of the speech rate), result in a signal with a quasi-periodic amplitude envelope.

If these assumptions hold, then at any point in time, it is possible to specify a measure of syllabic phase  $\phi$  (an angle), which varies continuously, and is dependent upon the quasi-periodic changes in the envelope of speech at or near to a characteristic frequency determined by speech rate. The value of syllabic phase associated with each part of the syllable is arbitrary, but for convenience we can specify a measure such that for energy maxima  $\phi \approx 0.0$ , whereas for energy minima,  $\phi \approx \pm\pi$  (given the above assumptions about the signal).

The phase information is yielded by convolving the signal with a temporal quadra-



ture filter. Filters like these have been used to determine local velocity in models of low-level visual motion processing (see e.g. Fleet & Jepson, 1990; Fleet & Langley, 1995); A quadrature filter (by definition) has two phase-shifted components. One component (the real one) is most responsive to periodic variations in amplitude in-phase with a cosine wave of a particular frequency determined by the expected periodicity of the input. The other component (the imaginary one) is most responsive to amplitude variations in-phase with a sine wave (with the same frequency). In combination, the output from the two components specifies a unique phase for any approximately quasi-periodic signal whose frequency is close to the tuning frequency.

For psychological plausibility, it is important that the filter makes use of only past and present information (i.e., it is causal). Furthermore, it should not require the storage of input over an arbitrary duration. Therefore, the degree to which output at a given time is affected by a prior input value decreases (exponentially) as their temporal separation increases. This is achieved by applying a (temporally-truncated) exponential envelope to the periodic components of the filter. Similarly shaped ('gammatone') filters have been used to model the early auditory processing in the middle ear (e.g. Patterson, Allerhand, & Giguere, 1995). Here, however, the proposed filter is tuned to a frequency well below the threshold of human hearing, but is instead sensitive to quasi-rhythmic changes in the envelope of the speech waveform. Such filters can be used to provide useful information about the syllabic and subsyllabic structure of an utterance.

The proposed filter can be described by a complex function of time with which the input time series is effectively convolved at each timestep. Before discussing the performance of the algorithm as applied to real speech, it may be useful to show how it works for an idealized signal (one that fulfils the assumptions set out above). The process is illustrated in figure 1.2. The simplest possible example is a sinusoid of the same frequency as the sinusoidal components of the filter. At each point in

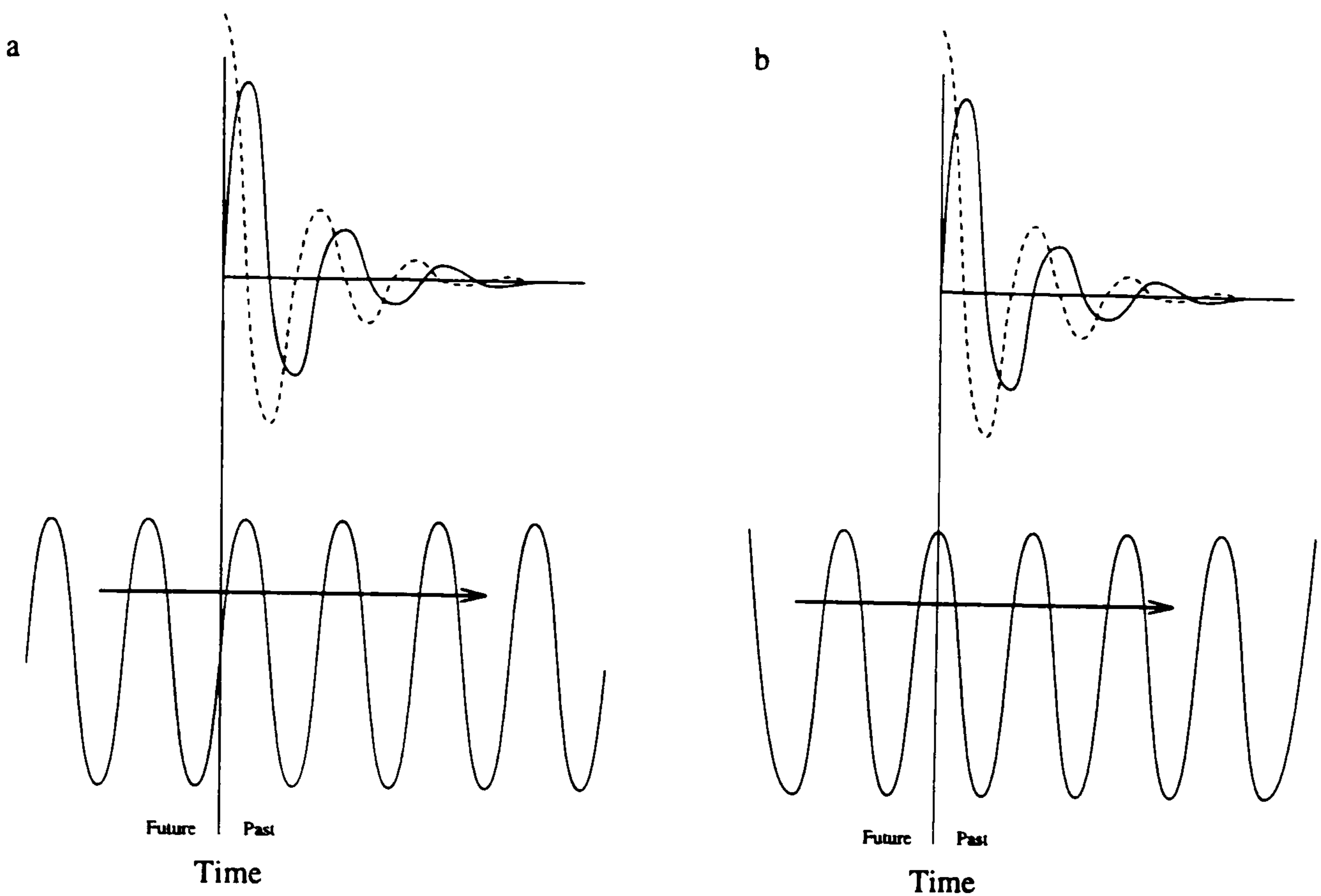


Figure D.4: A schematic representation of the temporal filtering of an ideal signal (a periodic one with frequency  $f$ ). The quadrature filter (represented on the upper axes) can be used to recover a phase description

time, the signal is convolved with the filter described above. The input signal can be regarded as moving (parallel to the time axis) relative to the stationary temporal filter. When the peaks in the signal are aligned with the peaks in the real component of the filter, the real component of the filter's output is strongly positive. Because the imaginary component of the filter has a phase which is shifted by  $\pi/2$  relative to the real part, the positive and negative elements in the convolution cancel out, and the output of the imaginary component is zero. Similarly, when the signal is in-phase with the imaginary component of the filter, its output is positive, while the real part's is zero.

As time progresses the output of the two filters vary such that they trace out a circular path in the complex plane (figure 1.2). The phase of the filter's output is the angle ( $\phi$ ) that the output makes with the origin. Since there exists such an angle for any non-zero complex number, the filter will yield a value of  $\phi$  for any non-zero

## 1. DYNAMICS OF THE SYLLABIC TEMPLATE

---

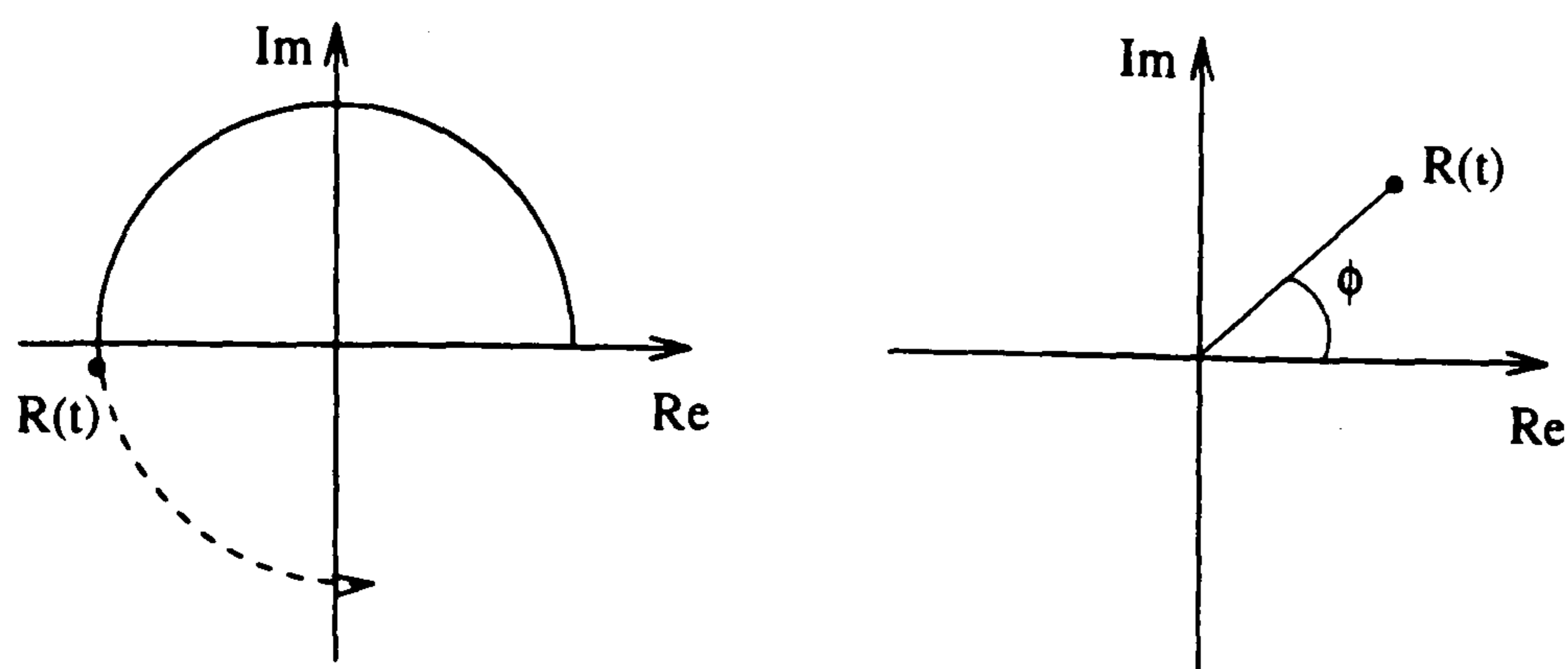


Figure D.5: *For an ideal signal, the output of the filter ( $R$ ) traces a circular path in the complex plane. The phase of the signal ( $\phi$ ) is the angle  $R$  makes with the real axis at any time.*

input, regardless of its frequency content. Moreover, the phase of the filter's output will not be sensitive to the overall amplitude of the input, only to local changes in amplitude at, or near to the central frequency. Quadrature filters do not have to be precisely tuned to the input signal to behave in this way; they can track phase accurately over a range of frequencies close to their central tuning (Fleet & Jepson, 1993), so that providing syllables are produced fairly rhythmically (assumption 2 above), the phase response of the filter should correspond to syllabic phase.

The convolution of the timeseries with the temporal filter would appear to demand the explicit buffering of acoustic information. In fact, though, it is not necessary to store previous input values explicitly – the filter can be implemented recursively (as an Infinite Impulse Response Filter). This means that the filter is capable of operating in real time on a dynamic input of arbitrary length, without buffering the input – previous input values are captured implicitly by feedback in the filter itself. The formal specification of the algorithm used is given in section 2.

In order to test the viability of the proposed algorithm, an implementation was developed to process sampled speech<sup>2</sup>. The speech sample takes the form of a series of signed integers representing the amplitude of the speech signal over a short interval,

---

<sup>2</sup>The SFS package from UCL phonetics department was used to store and display the data.

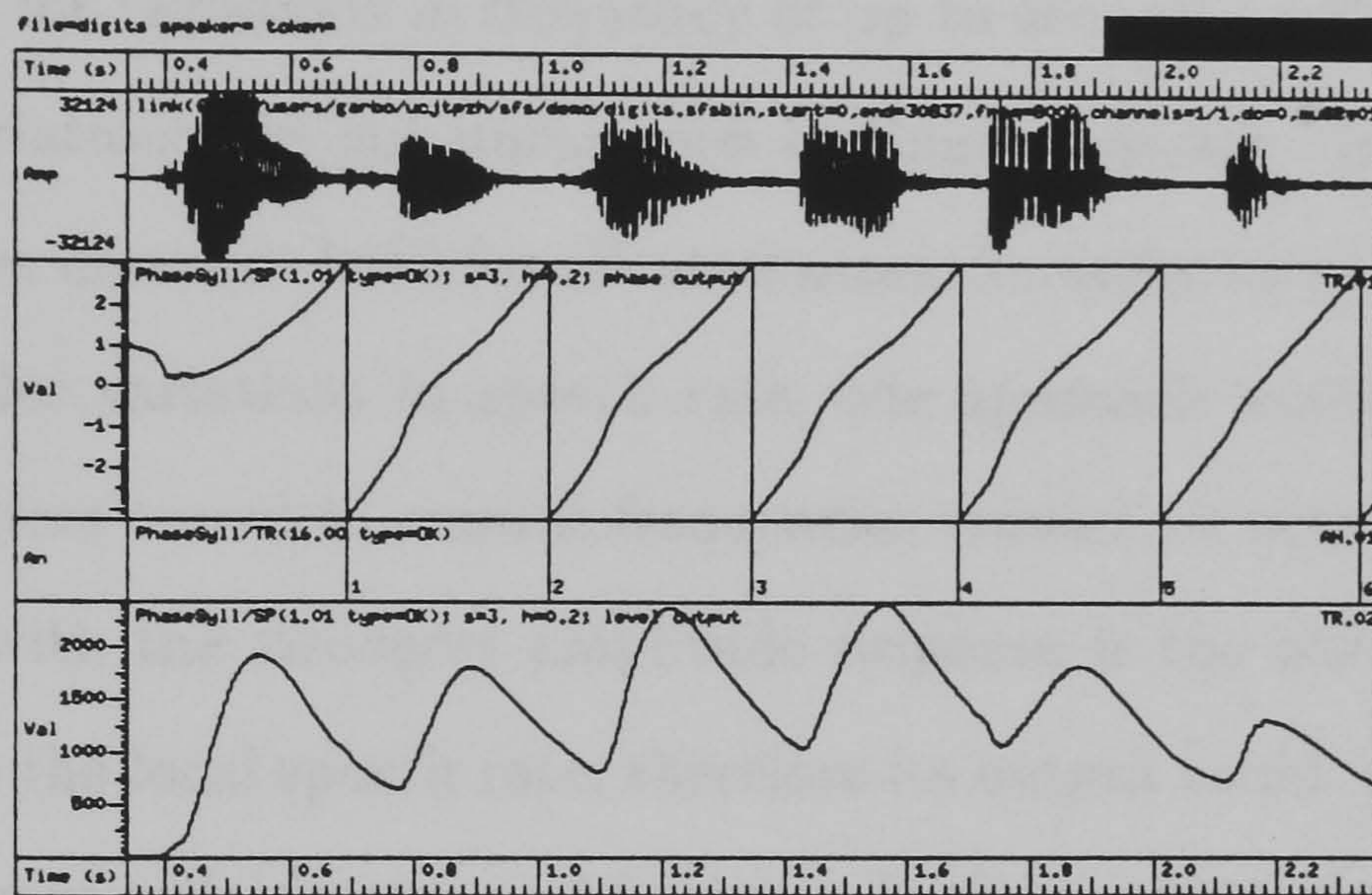
determined by the sampling rate of the hardware used. Because the proposed algorithm makes use of very low-frequency (below 10 Hz) information about the envelope of the signal, it is not necessary to use a particularly high-fidelity representation of the speech waveform. In fact, the audio hardware from Sun SPARCstation ELC was used to record the 'stimuli'. The supplied condenser microphone was used to record samples at a rate of 8000 Hz, to provide a digital recording of telephonic quality. The logarithmically compressed sampled waveforms were converted to linear pulse code modulation (PCM) samples before analysis.

The proposed filter requires as its input a representation of the envelope of the signal. In practice this can easily be achieved by rectifying the signal. It is also usual, when one is concerned with the envelope, to smooth the rectified signal. However, in this case smoothing is unnecessary, since the filter has an inherent bandpass character, which will prevent high-frequency components of the signal affecting its output. Therefore the absolute value of successive frames of the sampled waveform was used as the input to the filter described in equations 1 and 2 in section 2 below, and the changing value of  $\phi$  (equation 3) was recorded. Theoretical syllable boundaries can be identified by determining the times at which  $\phi = \pm\pi$ , and these points were determined by interpolation of the phase output of the filter.

Figure D.6 shows an example of the time course of syllabic phase for some sampled speech. The recording being analysed is of the author counting "one, two, three, four, five...". Each syllable corresponds to one cycle of the syllabic phase measure lower axes. Also shown are the points where the algorithm detects a syllable boundary (these are determined by interpolation of the syllabic phase measure). It is apparent that for this stimulus the syllabic phase measure tracks syllable structure as intended, with syllable boundaries occurring between vocalic peaks. Note that because each phoneme lasts for many timesteps, and has no clear start and end points, there is no direct correspondance between syllabic phase and phoneme identity, so

# 1. DYNAMICS OF THE SYLLABIC TEMPLATE

---



w a n t u θ r i f ɔ f e i v s i k s

Figure D.6: Output from an implementation of the syllabic phase filter. The top graph shows the speech waveform. Beneath it is the phase output of the filter, giving a real-valued measure of within-syllable position for each frame of the sample. Beneath the phase output, the theoretical syllable boundaries are marked. The bottom graph shows the amplitude of the filter's output. Beneath the graphs an approximate phonetic transcription is shown.

it is quite possible for ambisyllabicity to arise, for example an /n/ sound between two vowels might span the syllable boundary between them.

Although these results are promising, further work on the phase algorithm is necessary. Performance is very good for speech where the rate of production is fairly even, as it is here. However, initial experimentation shows that while the filter is tolerant of quite large local variations in speech rate (theoretically, phase information should be stable for variations in frequency of up to around  $\pm 50\%$  Fleet & Jepson, 1993), larger variations are not uncommon in English speech. In other words assumption 2 above does not hold for all utterances. In order to make the algorithm more robust under variations in speech rate, one approach would be to use a set of quadrature filters tuned to central frequencies around an octave apart. At any time, the filter with the strongest amplitude response is the one with the central tuning closest to the local speech rate, therefore its output could be used to give an accurate measure of syllabic phase. Alternatively the outputs of an array of filters could be arithmetically combined (by vector addition).

## 2 Formal specification of the syllabic phase algorithm

This section gives a formal description of the “syllabic phase” algorithm described above. Although at a conceptual level, syllabic phase is determined by convolution of the speech envelope with a low-frequency temporal quadrature filter, a recursive implementation has been employed. This implementation is based on work by Fleet and Langley (1995) and Clifford, Langley, and Fleet (1995) involving similar filters in models of motion processing in vision. The advantage of using a recursive implementation is essentially that information about prior inputs is stored implicitly –

## 2. FORMAL SPECIFICATION OF THE SYLLABIC PHASE ALGORITHM

---

the input history of the filter is captured by feedback in the system. A non-recursive implementation would require that previous timesteps be represented explicitly, in a spatially-distributed fashion, presumably in some kind of shift-register of finite capacity. This is an approach to the representation of temporally-distributed information, also taken by some connectionist models (see chapter 2), which I wish to avoid. The implementation described here, requires a very small amount of explicit storage (two-timesteps).

Equations for the real and imaginary components of the filter's output ( $Re[R]$  and  $Im[R]$ ) are given separately below. The recursive implementation requires that an intermediate complex value,  $Y$ , is first calculated and stored from one time step to the next. Throughout, the central frequency of the filter is expressed as an angular frequency,  $\omega = f/2\pi = 2\pi/r$ , where  $f$  is the central frequency of the filter. The other parameter, which determines the shape of the exponential envelope, is expressed as a decay constant  $b = \ln(2)/hs$  where  $h$  is the halflife, and  $s$  is the sample rate.  $I(t)$  is the input at time  $t$ .  $Y$  is given by:

$$\begin{aligned} Re[Y(t)] &= I(t) - Re[Y(t-1)] \frac{b^2 - 4}{(b+2)^2 + \omega^2} - Im[Y(t-1)] \frac{4\omega}{(b+2)^2 + \omega^2} \\ Im[Y(t)] &= -Im[Y(t-1)] \frac{b^2 - 4}{(b+2)^2 + \omega^2} + Re[Y(t-1)] \frac{4\omega}{(b+2)^2 + \omega^2} \end{aligned} \tag{1}$$

The complex output of the filter  $R$  is calculated using the present and previous values of  $Y$ :

$$\begin{aligned}
 \text{Re}[R(t)] &= (\text{Re}[Y(t)] + \text{Re}[Y(t-1)]) \frac{b^2 + 2b}{(b+2)^2 + \omega^2} - \\
 &\quad (\text{Im}[Y(t)] + \text{Im}[Y(t-1)]) \frac{2\omega b}{(b+2)^2 + \omega^2} \\
 \text{Im}[R(t)] &= (\text{Im}[Y(t)] + \text{Im}[Y(t-1)]) \frac{b^2 + 2b}{(b+2)^2 + \omega^2} + \\
 &\quad (\text{Re}[Y(t)] + \text{Re}[Y(t-1)]) \frac{2\omega b}{(b+2)^2 + \omega^2}
 \end{aligned} \tag{2}$$

Syllabic phase,  $\phi$  is the inverse tangent of the ratio of imaginary and real components of the filter's output:

$$\phi = \arctan \left( \frac{\text{Im}[R(t)]}{\text{Re}[R(t)]} \right) \tag{3}$$

The amplitude of the filter's output is simply  $|R| = \sqrt{\text{Im}[R(t)]^2 + \text{Re}[R(t)]^2}$

The phase response of the filter is linear for periodic signals of frequency  $f$ . It degrades gracefully as the frequency of the input signal moves away from its tuning (Fleet & Jepson, 1993). Phase is not sensitive to the size of the amplitude variations in the signal, since it is a function of the ratio of the real and imaginary components of the filter's output (equation 3). Two parameters must be specified, one,  $h$ , determining the decay of temporal information in the filter and another,  $f$ , determining the frequency to which the filter is tuned. The optimal value of  $f$  is determined by the speech rate (in syllables per second);  $h$  must be large enough to allow the filter to make use of information about the temporal context of the input. Trial and error suggests that decay should have a half-life not much less than the duration of one syllable. From a psychological point of view, however it should not be so small that it suggests the implicit storage of low level acoustic information for much more than a few seconds.



# Bibliography

- Amit, D. J., Sagi, D., & Usher, M. (1990). Architecture of attractor neural networks performing cognitive fast scanning. *Network, 1*, 189–216.
- Atkinson, R. C., & Shiffrin, R. M. (1968). The control of short-term memory. *Scientific American, 225*, 82–90.
- Baars, B. J. (1976). The competing plans hypothesis: a heuristic viewpoint on the problem of speech errors.. In *UCLA Working Papers in Linguistics*.
- Baars, B. J., & Motley, M. T. (1974). Spoonerisms: experimental elicitation of human speech errors. *JSAS Catalog of selected documents in psychology, ?*, ?
- Baars, B. J., & Motley, M. T. (1976). Spoonerisms as evidence of sequencer conflicts: evidence from artificially elicited errors. *American Journal of Psychology, 89*(3), 467–484.
- Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362–365.
- Baddeley, A. D. (1986). *Working Memory*. Clarendon, Oxford.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In Bower, G. (Ed.), *The psychology of learning and motivation*. Associated Press, New York.

## BIBLIOGRAPHY

---

- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586–596.
- Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, *9*, 176–189.
- Basso, A., Spinnler, H., Vallar, G., & Zanobio, E. (1982). Left hemisphere damage and selective impairment of auditory-verbal short-term memory: a case study. *Neuropsychologia*, *20*, 263–74.
- Bisiacchi, P. S., Cipolotti, 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.
- Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behaviour*, *13*, 80–97.
- Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, *3*, 2–12.
- Burgess, N. (1995). A solvable connectionist model of immediate recall of ordered lists. In Tesauro, G., Touretzky, D., & Leen, T. K. (Eds.), *Advances Neural Information Processing Systems*, *7*. MIT.
- Burgess, N., & Hitch, G. (1992). Towards a network model of the articulatory loop. *Journal of Memory and Language*, *31*(4), 429–460.
- Butterworth, B. (1992). Disorders of phonological encoding. *Cognition*, *42*, 261–286.
- Campbell, R., & Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits

## BIBLIOGRAPHY

---

- 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.
- Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: evidence from a case of acquired dyslexia. *Cognition*, 26(?), 59-85.
- Chi, M. T. H. (1976). Short-term memory limitations in children: capacity or processing deficits. *Memory and Cognition*, 4, 559-572.
- Clements, G. N., & Keyser, S. J. (1981). A three-tiered theory of the syllable. Tech. rep., Massachusetts Institute of Technology.
- Clifford, C. W. G., Langley, K., & Fleet, D. J. (1995). Centre-frequency adaptive IIR temporal filters for phase-based image velocity estimation. In *Image Processing and Its Applications* Heriot-Watt University, UK.
- Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology*, 55, 429-432.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.
- Cutler, A. (1982). The reliability of speech error data. In Cutler, A. (Ed.), *Slips of the tongue and language production*, pp. 7-28. Mouton, Hague.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.

## BIBLIOGRAPHY

---

- Dell, G. S. (1988). The retrieval of phonological forms in production: tests 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.
- Ebbinghaus, H. (1964). *Memory: a contribution to experimental psychology*. Dover.
- 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.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Ericsson, K. A., Chase, W. G., & Falloon, S. (1980). Aquisition of a memory skill. *Science*, 208, 1181–2.
- Estes, W. K. (1972). An associative basis for coding and organisation in memory. In *Coding processes in Human memory*, pp. 161–190. Winson, Washington, DC.
- Fleet, D. J., & Jepson, A. D. (1990). Computation of component velocity from local phase information. *International Journal of Computer Vision*, i5, 77–101.
- Fleet, D. J., & Jepson, A. D. (1993). Stability of phase information. *IEEE Pattern Analysis and Machine Intelligence*, 15, 1253–1268.
- Fleet, D. J., & Langley, K. (1995). Recursive filters for optical flow. *IEEE PAMI*, 17, 61–67.
- Freud, S. (1901). *Psychopathology of everyday life*. Benn, London. Translated by A. Tyson, 1966.
- Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5, 193–320.

## BIBLIOGRAPHY

---

- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. In *Slips of the tongue and language production*, pp. 251–263. Mouton.
- 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., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language*, 29, 336–360.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. 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 Levy, J., Bairaktaris, D., Bullinaria, J., & Cairns, D. (Eds.), *Connectionist models of memory and language*. UCL Press.
- Goldstein, L. (1977). Categorical features in speech perception and production.. Paper presented to the working group on slips of the tongue chaired by V. Fromkin; XII International Congress of Linguists; Vienna, Austria.
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Erlbaum, London.

## BIBLIOGRAPHY

---

- Grossberg, S. (1978). A theory of human memory: Self-organisation and performance of sensory motor codes, maps and plans. In Rosen, R., & Snell, F. (Eds.), *Progress in Theoretical Biology*. Academic Press, New York.
- Guildford, J. P., & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology*, 36, 621–628.
- Henson, R., Norris, D., Page, M., & Baddeley, A. (1995). Unchained memory: error patterns rule out chaining models of immediate serial recall..
- Houghton, G. (1990). The problem of serial order: a neural network model of sequence learning and recall. In Dale, R., Mellish, C., & Zock, M. (Eds.), *Current Research in Natural Language Generation*. Academic Press, London.
- Houghton, G., Glasspool, D. W., & Shallice, T. (1994). Spelling and serial recall: insights from a competitive queueing model. In Brown, G. D. A., & Ellis, N. C. (Eds.), *Handbook of normal and disturbed spelling*. Wiley, Chichester.
- Houghton, G., & Hartley, T. Parallel models of serial behaviour: Lashley revisited..
- 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.
- Jordan, M. I. (1986). Serial order: a parallel distributed processing approach. Tech. rep. 8604, Institute for Cognitive Science, University of California, San Diego.
- Ladefoged, P. (1993). *A course in phonetics (third edition)*. Harcourt Brace College Publishers, Fort Worth.
- Lashley, K. S. (1951). The problem of serial order in behaviour. In Jeffress, L. A. (Ed.), *Cerebral mechanisms in behavior*. Wiley, New York.

## BIBLIOGRAPHY

---

- Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, 96, 25-27.
- Mackay, D. G. (1972). The structure of words and syllables :evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.
- MacKay, D. G. (1987). *The Organisation of Perception and Action*. Springer-Verlag, New York.
- Martin, N., Saffran, E. M., Dell, G. S., & Schwartz, M. F. (1994). Origins of parasphasias in deep dyslexia: testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval..
- Mechanic, A. (1964). The responses involved in the rote learning of verbal materials. *Journal of Verbal Learning and Verbal Behaviour*, 3, 30-36.
- Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 2, 1-21.
- Meringer, R., & Mayer, K. (1895). *Versprechen und verlesen: eine psychologisch-linguistische Studie*. Göschel, Stuttgart.
- 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.
- Miller, G. A. (1956). The magical number seven, plus or minus two:some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Motley, M. T. (1973). An analysis of spoonerisms as psycholinguistic phenomena. *Speech Monographs*, 40, 66-71.
- Motley, M. T., & Baars, M. J. (1975). Encoding sensitivity to phonological markedness and transitional probability. *Human Communication Research*, 2(1), 353-371.

## BIBLIOGRAPHY

---

- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 21, 263–276.
- Nooteboom, S. G. (1969). The tongue slips into patterns. In van Essen, A. J., & van Raad, A. A. (Eds.), *Leyden studies in linguistics and phonetics*, pp. 114–132. Mouton, Hague.
- 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.
- Patterson, R. D., Allerhand, M., & Giguere (1995). Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–344.
- Quinlan, P. T. (1992). *The Oxford Psycholinguistic Database*.
- Rumelhart, D. E., Hinton, G., & Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., & McClelland, J. L. (Eds.), *Parallel distributed processing, volume 1: foundations*. MIT press, Boston, Mass.
- Rumelhart, D. E., & Norman, D. (1982). Simulating a skilled typist: a study of skilled motor performance. *Cognitive Science*, 6, 1–36.
- Selkirk, E. (1984). On the major class features and syllable theory. In Aronoff, M., & Oehrle, R. T. (Eds.), *Language sound structure: studies in phonology presented to Morris Halle by his teachers and students*. MIT press.



## BIBLIOGRAPHY

---

- Shaffer, L. H. (1976). Intention and performance. *Psychological Review*, 83(5), 375-393.
- Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, 15, 729-735.
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: a neuro psychological case study. *Quarterly Journal of Experimental Psychology*, 22, 261-273.
- Shattuck, S. R. (1975). *Speech errors and sentence production*. Ph.D. thesis, MIT.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In Cooper, W. E., & Walker, E. C. T. (Eds.), *Sentence processing :psycholinguistic studies presented to Merrill Garrett*. Erlbaum, Hillsdale, N.J.
- Stemberger, J. P. (1985). An interactive activation model of language production. In Ellis, A. (Ed.), *Progress in the Psychology of Language*, pp. 143-186. Erlbaum, Hillsdale, N.J.
- Svartvik, J., & Quirk, R. (1980). *A corpus of english conversation*. Gleerup.
- Tipper, S. P., Howard, L. A., & Jackson, S. R. Selective reaching to grasp: evidence for distractor effects..
- Treiman, R. 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.
-

## BIBLIOGRAPHY

---

- Treiman, R., Straub, K., & Lavery, P. (1994). Syllabification of bisyllabic nonwords: evidence from short-term memory errors. *Language and Speech, 37*(1), 45–60.
- Vallar, G., & Baddley, A. D. (1982). Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology, 34*, 53–60.
- Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78*, 567–574.
- Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behaviour. *Psychological Review, 76*, 1–15.
- Wickelgren, W. A. (1976). Phonetic code and serial order. In Carterette, E. C., & Friedman, M. P. (Eds.), *Handbook of Perception*, Vol. 7, pp. 225–264. Academic Press.