Modelling serial order in behaviour: Studies of spelling

David William Glasspool

University College London

Thesis submitted for the degree of Doctor of Philosophy

March 1998

Abstract

Serial order in behaviour remains an interesting problem for computational modelling in psychology, especially for connectionist approaches. The 'Competitive Queuing' (CQ) approach to sequence generation has the advantage of accounting for a number of common features apparent in several different types of serial behaviour. This thesis addresses the general account which the CQ approach can give for constraints on serial errors within sequences by developing models of an acquired disorder of spelling, 'graphemic buffer disorder' (GBD). Two approaches to the development of a simple initial model of GBD into more complex models are demonstrated, and are related to the general problem of accounting for serial category constraints in sequencing.

The initial CQ model of GBD is based on an existing model of speech production with minimal spelling-specific changes. A number of shortcomings are identified in the performance of this model, in particular the inability to distinguish consonant and vowel letters, which prevents a striking feature of GBD errors - the preservation of consonant/vowel status - from being modelled. An analysis of the general problem of adding domain-specific constraints to CQ models suggests two approaches to improving the initial model. Two alternative extended models are thus advanced. The first is a development of the initial model incorporating an external template to specify consonant/vowel information. Simulations with this model demonstrate a much improved fit to the data. The second model develops a novel architecture, generalising the CQ approach to multi-layer networks. The model is less detailed but demonstrates the correct general features of the GBD error pattern. The relationship between the models is discussed and possible future research directions are identified.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank George Houghton, both for providing the original impetus to move to London to work in this field and for his supervision of this thesis. His help and encouragement, and his efforts to keep my work focused and productive, have been much appreciated.

I also thank Tim Shallice for many discussions about all aspects of this work, but especially the material in Chapters 3 and 7, which have benefited greatly from his comments and suggestions.

Finally, I am grateful to my wife, Heather, whose love, support and encouragement over many years, along with an amazing tolerance for the realities of academic life, have made this thesis possible. This thesis is dedicated to her, with my thanks.

Preface

Many types of behaviour have a temporal aspect - it matters not just what actions are taken, but in what order they are taken. This is true of everyday motor tasks such as getting dressed or driving a car, and is central to all types of linguistic behaviour. The need to deal with the serial ordering of responses is thus of great importance to much of psychology. Yet historically the issue of serial order has received relatively little attention from the cognitive sciences. In a seminal paper of 1951 Karl Lashley described what he termed "the problem of temporal integration" as "the most important and also the most neglected problem of cerebral physiology" (Lashley, 1951, p.508). Lashley's paper has become a rallying point for subsequent work in the area, and the situation is considerably better now than he found it in 1951, with a two-pronged attack currently being mounted by experimental studies and the development of more realistic computational models of serial behaviour. Perhaps a sign that the field is beginning to reach maturity is the two-way interaction which is becoming evident between the two approaches, with modelling work now starting to inform empirical studies (see for example Henson and Burgess, 1997, Henson, submitted, Henson, Burgess and Frith, in preparation).

The general questions for computational approaches to the problem of serial order are: What type of mechanisms underlie the ability to generate sequences of actions, and what are the relationships between the mechanisms of serial order in different types of behaviour? These are broad questions which will not soon be answered, although some general principles may already be discerned in the experimental and modelling literature. This thesis concerns a particular approach to the generation of serial behaviour, the 'Competitive Queuing' approach, and proceeds through the development of a number of models of an acquired disorder of spelling - Graphemic Buffer Disorder. By refining the account which Competitive Queuing can give for serial behaviour the thesis aims to contribute towards answering these questions.

Spelling might seem at first to be an odd choice of domain, being an acquired skill of relatively recent origin and thus apparently somewhat artificial. In fact this very artificiality is one of the attractions of spelling as a target for serial modelling. Other

linguistic skills such as speech production or verbal short-term memory no doubt tap highly domain-specific mechanisms which have evolved specifically to handle the idiosyncrasies of verbal language, whereas spelling, in the absence of any specially evolved mechanisms, may rely more on general purpose underlying serial capabilities. The particular spelling disorder considered in this thesis is limited mainly to a problem with the sequencing of responses, independent of high-level linguistic structure, and thus holds the hope of relatively direct access to the underlying mechanisms of sequence production.

Chapter 1 provides a general introduction to the thesis, intended to situate the modelling work of later chapters within the wider field of serial behaviour. The chapter reviews experimental studies across several domains with the aim of showing that a number of common features are evident in serial behaviour within apparently quite different areas. This observation has two important consequences. Firstly, it allows the utility of spelling as a domain within which to develop general serial models to be assessed - the models will be generally applicable to the extent that spelling is a typical form of serial behaviour. Secondly, the existence of common features in the experimental data is important in informing the choice of modelling paradigm adopted - modelling frameworks gain importance to the field of serial behaviour to the extent that they can explain why these common features should occur. Chapter 2 reviews the approaches which have been taken to generating serial behaviour in computational models and concludes that the Competitive Queuing (CQ) approach is particularly promising in this regard. Chapter 3 then introduces the particular spelling disorder which the modelling work will address - Graphemic Buffer Disorder (GBD). This pathology promises to be a useful testing ground for serial models because it is essentially a disorder of the sequencing of letters. In Chapter 4 a basic model of GBD spelling is developed within the CQ framework - the CQS model. The model's performance is assessed and desiderata for a more detailed model are established. Chief amongst these is the need to represent more than just the basic sequence of letters making up a word, since a feature of GBD is that the consonant or vowel status of letters is often preserved when spelling errors occur. Refining the model to account for this feature raises the general issue of domain-specific constraints in serial behaviour,

and Chapter 5 develops a framework for the implementation of such constraints in CQ models. The framework suggests two possible directions for the further development of the GBD model. Chapter 6 takes the first of these possibilities as the basis for a comprehensive development of the CQS model to produce a more detailed account of GBD. Chapter 7 uses the second of the possibilities as the impetus for a different approach to the problem, generalising the CQ framework to multi-layer networks. This final model is a relatively general one which may provide the basis for more detailed models in other domains as well as GBD. Finally, Chapter 8 discusses the models in more general terms and draws some overall conclusions.

1

Serial order in behaviour

Serial order in human behaviour has been extensively studied in three major areas: language, short-term memory and motor behaviour. This chapter will review the relevant data in these areas and draw some parallels between a number of the findings from disparate studies.

There are three major types of evidence which may be used in the experimental determination of the structure underlying an intact psychological competence: Timing information and patterns of errors in normal subjects, and the patterns of breakdown of the normal competence observed in neurological patients. Given the simple existence of a psychological competence - say, the ability to write a word from memory - there are very few limits on the types of mechanism which might be proposed to account for it. For example, any number of computer programs can be written which will produce a sequence of letters correctly. The higher informational content provided by timing, error and breakdown data reduces the range of possible mechanisms which might underlie the human competence. Timing information and error patterns may be collected from normal subjects and are thus very useful in discriminating between alternative putative mechanisms. Both have been elicited in numerous experimental studies. The two types of data have different typical uses. Timing data can give an indication that something different is happening at a particular point in a process, and this has been used, for example, in studies of hierarchical chunking in the motor domain. Errors can carry a great deal of information about the ways in which the system can break down, and have been used extensively to direct theories of the mechanisms which generate and control serial behaviour. Many aspects of the subject's errors may be used. In addition to a crude count of errors, their position, the way they are influenced by load and data type, and their implications concerning what information is available at any point during response production may all be useful in driving theory. The breakdown of normal behaviour in neuropsychological syndromes places more general, structural constraints on models, although the detailed structure of

errors made by patients may also constrain the internal mechanisms posited within processing 'modules' (Shallice, Glasspool and Houghton, 1995).

1.1 Language

Language is serially structured behaviour *par excellence*. A large part of the informational content of language is carried in the serial order of its constituents, and generating, controlling and extracting that order information is an important aspect of linguistic communication. Language, then, should be a major source of evidence for the operation of serial systems in the brain. The processes involved in both major modes of language production, speaking and writing, have been subject to extensive experimental investigation.

1.1.1 Speech

The main source of information on the mechanisms underlying speech production is the analysis of speech errors. Additional information comes from the study of two specific phenomena, co-articulation and 'tip-of-the-tongue' states.

Speech Errors

Many studies have identified patterns in collections of speech errors (e.g. Shattuck-Hufnagel, 1979, Baars and Motley, 1976). Dell (1986) and Levelt (1989) both review the findings of a number of studies. The following main findings have emerged as a general consensus:

- Erroneous items in an utterance may be displaced forwards or backwards from their correct position by some distance, often over several intervening items. Lashley (1951) took this to indicate that there is a level in the speech production process at which an extended series of sounds are concurrently present.

- All of the five possible types of error which can be made in the production of any serially ordered sequence of items, Exchanges (ABCD → DBCA), shifts (ABCD → BCDA), insertions (ABCD → ABCAD), deletions (ABCD → ABD) and

substitutions (ABCD → ABCA), can be identified in speech errors. All occur within each linguistic level or category, e.g. phonemes, morphemes, syllables or words (Shattuck-Hufnagel, 1979).

- The 'chunk' of an utterance which is involved in an error generally corresponds to a unit which can be identified as a single segment according to linguistic theory (for example, a single phoneme, morpheme, word or phrase) (Fromkin, 1971, 1973, 1980). Moreover, when two items interact in an error (as when two items are transposed, for example, or when one item is substituted for another) the interacting items are nearly always units at the same linguistic level or members of the same linguistic category (Fay and Cutler, 1977).

- The interacting items in errors usually share characteristics. Thus two exchanged phonemes will often share phonemic features, exchanged words often share the same syntactic class, and so on (Shattuck-Hufnagel, 1979). Errors involving the exchange of phonemes between words almost invariably involve phonemes from the same syllabic position (Motley, 1973).

- Repeats of particular items or similar segments facilitate errors. Dell (1984), for example, found that repeated phonemes in speech lead to an increase in order errors.

Timing studies in speech

The main finding of studies of reaction time in speech is that time taken to begin an utterance increases with its length (Sternberg, Monsell, Knoll and Wright, 1978). A similar effect is also found for typing. This has been interpreted as indicating that time is required to assemble an utterance in some form of buffer prior to the start of speaking (Rosenbaum, 1991).

Co-articulation

The phenomenon of co-articulation (Moll and Daniloff, 1971, Benguerel and Cowan, 1974) may also throw light on the mechanisms of speech production. Subject to physical constraints the vocal tract anticipates up-coming phonemes during articulation, so that the current phoneme tends to take on some of the features of to-be-articulated

phonemes. Co-articulation is not simply the result of interaction within physiological mechanisms, as demonstrated by the fact that it is language dependent - which phonemic features may be changed to allow co-articulation on a particular phoneme depends on the norms of the target language (Jordan, 1986). Co-articulation is further evidence for the proposition that phonemes are not articulated by totally serial processes, but that at some level a number of phonemes are represented in parallel.

Tip-Of-The-Tongue States

A tip-of-the-tongue (TOT) state is one in which a word which is known cannot be recalled, though the subject experiences a strong sense that recall is imminent and that the word is 'just out of reach' or 'on the tip of their tongue'. Such experiences appear to be fairly common, occurring around once a week in daily life (Brown, 1991), and are intriguing in that the subject often has access to aspects of the structure of the target word. Brown (1991) reviews the findings from several studies and finds the following features:

- While in a TOT state, subjects often retrieve what appear to be fragments of the target or related words, or structural aspects of the target. Brown and McNeill (1966) found that when this happened 70% of fragments came from similar sounding related words and 30% from words with similar meanings.

- The first letter and the first phoneme letter of the target word both seem to be accessible to a degree significantly higher than chance - Brown (1991) reports 50% to 70% correct first letter reports across several investigations, against a baseline of around 10% for guesses with words the subjects did not know. A "U"-shaped curve for probability of correct report against position of letter in word has been found by more than one investigation (e.g. Tweney, Tkacz and Zaruba, 1975). However, Koriat and Lieblich (1975) suggest this may be a statistical artefact, since words in general are more similar in their final segments than in their initial segments.

- The number of syllables in the target word is particularly likely to be accessible in a TOT state. Brown and McNeill (1966) find a success rate of 60%, Koriat and Lieblich (1974) report 80%.

1.1.2 Writing

Writing is a relative latecomer among linguistic skills. It is unlikely, therefore, that any brain mechanisms have specifically evolved to support writing in the way that special purpose mechanisms presumably support speech. Nonetheless, neuropsychological evidence shows that once written language has been acquired it is localised in the brain in the same manner as spoken language, and several well established neuropsychological syndromes exist in which written language is compromised in a systematic manner. In the case of writing most experimental work on the underlying mechanisms has focused on two aspects: spelling and rapid skilled typing.

Spelling - Normal Spellers

In studies of spelling it is the subjects' spelling errors rather than their reaction times which have been examined. Wing and Baddeley (1980) investigated spelling slips (rather than misspellings through not knowing the correct spelling) in normal subjects by analysing a corpus of errors from Cambridge University entrance examination papers. They found the following error types:

- Omissions   (49%)   E.g. likely → likly

- Substitutions   (36%)   E.g. *desirable* → desireble

- Insertions   (13%)   E.g. *political* → polictical

- Exchanges   (2%)   E.g. *cannot* → cannto

Wing and Baddeley found a clear 'inverted U' serial position curve, with most errors occurring in word-medial positions. Errors occurred no more often in final than in initial positions. They also found an effect of word length: Words with errors were reliably longer on average than those without.

Hotoph (1980) performed a similar analysis, with findings in line with those of Wing and Baddeley (although a somewhat different scoring system was used), with an 'inverted-U' serial error incidence curve and a similar distribution of different error types. Hotoph also finds an interesting non-significant tendency for errors to occur more frequently in words with repeated letters.

Spelling - Disordered spelling

A well-studied pathology of spelling of particular interest for this thesis is so-called Graphemic Buffer Disorder (GBD). This disorder is interesting because it appears to be primarily a deficit in the serial production of letters, rather than in access to abstract knowledge of spellings. Chapter 3 will review the disorder in some detail, but for now it may be noted that errors typically involve misorderings and substitutions of letters in target words. The errors have the following characteristics:

1. **Error types**: Errors of substitution, shift, exchange, insertion and deletion all occur, though with different frequencies.

2. **Word length effects**: The likelihood of an error increases with number of letters in the target word.

3. **Serial position effects**: Error rates increase towards the middle of words.

4. **Doubling errors**: Words containing a doubled letter show errors in which the property of doubling seems to dissociate from the letter being doubled (e.g. *school* → *schhol*).

5. **Preservation of consonant/vowel status**: The large majority of errors preserve the consonant or vowel status of the target letter.

Typing - Timing studies

Response timing studies in typing have primarily addressed the question of the underlying representation of the sequence of keystrokes required to type a word - specifically, whether typing is controlled by a *chain of associations* such that each keystroke is the stimulus for the next, or by some *motor plan* which is specifies the sequence of keystrokes in full before typing a word commences. Skilled typists routinely produce keystrokes at rates of up to 450 per minute (Rosenbaum, 1991), or one every 133 msec. The time taken to respond to external feedback is generally agreed to be in excess of 200 msec, which indicates that keystrokes are produced in accordance with a pre-determined internal motor plan. Incorrect keystrokes have longer

latencies than correct ones (Shaffer, 1975), and are made with lighter force (Rabbitt, 1978), indicating that they are detected before their movement has been completed. This suggests that although response chaining is unlikely, at some level the activity is being monitored at a high enough speed that some form of chaining based on internal processes cannot be ruled out on this evidence alone. However, the time taken to start typing a word increases with its length (Sternberg, Monsell, Knoll and Wright, 1978), which is consistent with the idea that a plan of some sort must be prepared before the typing of a word begins.

Typing - Error studies

Rumelhart and Norman (1982) obtained error and timing information during the typing to dictation of a 90000 keystroke manuscript by a single skilled typist, and subjected this corpus to a detailed error analysis. The major error types they discovered were:

- Transposition errors (*because* → beacuse).

- Omission errors (*amount* → amont).

- Homologous errors (the substitution of a letter hit by the corresponding finger on the opposite hand).

- Doubling errors. As with spelling errors of GBD patients, the property of doubling appears to dissociate from letter identity (e.g. *school* → scholl).

- Alternation reversal errors (*these* → thses). In a somewhat similar way doubling errors, the property of 'alternation' appears to be separable from the identities of the letters involved.

Most typing errors involve spatially close keys (Lessenberry, 1928), yet these are not due to the typist's finger slipping: the correct finger for the erroneous key is used, even when it is different to the finger required for the correct key (Grudin, 1983). This indicates that the error occurs prior to the level of processing at which finger movements are specified, and appears to reflect a spatial organisation of internal representations for keystrokes. Rumelhart and Norman (1982) also find 'co-

articulation' in typing, with the movement generated to make the current keystroke modified to take into account up-coming keystrokes.

1.2 Short-term Memory

Short-term memory for ordered lists of verbal items, for visual or spatial information and for sequences of movements have all been extensively investigated.

1.2.1 Verbal short-term memory

The main characteristics of verbal short-term memory (STM) are as follows:

- **Length effects.** These are of two kinds:

    1. *Sequence length (span):* The more words there are in a sequence, the poorer recall is, with near perfect performance up to around four words followed by a rapid cut-off to near zero performance at ten words (e.g. Guildford and Dallenbach, 1925, Baddeley, Thomson and Buchanan, 1975).

    2. *Word length:* capacity is greater for sequences of short words than for longer words. In fact, capacity appears to depend on the temporal length of the list to be recalled and the subject's speech rate rather than on the number of words or syllables in the list (Baddeley, Thompson and Buchanan, 1975, Hulme, Maughan and Brown, 1991, Ellis and Henneley, 1980) Subjects can recall about as many words as they can say in one to two seconds, plus a constant.

- **More order errors than item errors**: A higher proportion of errors involve the order of recalled items than involve the identity of the items. Aaronson (1968), for example, found that between 70% and 80% of errors involved the order of otherwise correctly recalled items.

- **Phonemic similarity effect**: Capacity for lists of words is impaired when the words are phonemically similar (e.g. Conrad and Hull, 1964, Baddeley, 1966).

- **Lexicality effect:** Recall is poorer for nonwords than for words (Besner and Davelaar, 1982, Watkins, 1977). Hulme, Maughan and Brown (1991) demonstrate that this difference is significantly greater than would be predicted by the slower speech rate of subjects for nonwords.

- **Serial error curves:** Recall is better for the first few items in the list (the primacy effect) and the last few items (the recency effect) (e.g. Murdock, 1962, Murray, 1966, Crowder, 1972).

- **'Chunking' effects:** Several studies find evidence for the movement of items either between lists in experiments in which multiple lists are presented consecutively, or between groups in experiments where subjects group items together to form sub-lists within the stimulus list. For example, Fuchs (1969) finds that intrusions from one list to a subsequent list in the same set of trials tended to occur to the same serial position. 'Chunking' of information into hierarchical structures is a useful strategy for improving recall (Miller, 1956; Simon, 1972). Simon (1974) draws common parameters from various past experimental results, and concludes that short-term and long-term memory are hierarchically organised, with the size of a 'chunk' in the hierarchy fixed for any individual.

- **Effect of repeats:** A number of studies have found worse recall of repeated items than (nonrepeated) control items at the same positions in control lists (the 'Von Ranschburg' effect, Crowder, 1968, Jahnke, 1969). Henson (in press, b) shows that this is true only when repeated items are separated by intervening items; recall is facilitated for repeated items when they are immediately adjacent.

It has often been claimed that the characteristic primacy and recency effects in verbal STM are evidence that at least two distinct mechanisms are involved, primacy being due to slow consolidation of early items in a long-term component, while recency is due to the contribution of a 'primary' or short-term component which is only able to store between two and four items (e.g. Atkinson and Shiffrin, 1968, Glanzer, 1972, Tzeng, 1973, Baddeley and Hitch, 1977). However, an alternative view is that serial position effects are due to different aspects of the operation of a unitary system. Greene (1986) reviews the evidence concerning the origin of the recency effect, which is often seen as

the strongest evidence for separate systems in STM, and concludes that the effect is best explained by 'end effects' in a unitary memory system using a positional coding scheme. Greene's main reasons for this view are as follows:

1. **The Continuous Distracter paradigm:** Normally a distracter activity immediately following presentation removes recency. (Glanzer and Cunitz, 1966, Postman and Phillips, 1965). This has been interpreted as disruption of a separate short-term store by the distracter activity. The Continuous Distracter paradigm (Bjork and Whitten, 1974) interposes a distracter task before each pair of items throughout presentation. In this case a final distracter period does not remove the recency effect.

2. **Concurrent Distraction:** Baddeley and Hitch (1977), among others, have presented stimulus lists alongside a concurrent distracter task. The distracter task is intended to fully occupy those processes responsible for rehearsal. According to the Articulatory Loop hypothesis (Baddeley and Hitch, 1974), preventing rehearsal in this way should disrupt the primary store, and hence recency. However, the recency effect is still present under such conditions.

3. **Multicategory lists:** Watkins and Peynircioglu (1983) presented stimulus lists containing inter-mingled sublists of distinct categories of items. Recall was required of one sublist at a time. A clear recency effect was found *within* sublists, a finding which is inconsistent with 'primary memory' interpretations of the recency effect.

1.2.2 Visual short-term memory

Hitch (1974, experiment 2) attempts to separate spatial from temporal order information in a short-term memory task. Letters were sequentially presented in random spatial positions in a horizontal array. Subjects were probed for the temporal and/or sequential position of a letter in the array. This procedure allowed separate serial error curves to be plotted for temporal and spatial probes. The serial position error curve for spatial probe shows no primacy effect, but a one-item recency effect.

Healy (1975) presented subjects with four successive consonants, each in a different spatial location. In the spatial order condition, the temporal order of the letters remained constant but their spatial order was varied. Healy found significant primacy and recency effects for this spatial condition, although their magnitude was less than those for a similar temporal order condition where the spatial order was held constant.

Neither of these studies is clearly independent of linguistic or motoric effects. However, both attempt to isolate the *retention* aspect of the procedure, which is limited to spatial information only. It is thus likely that any linguistic or motor processes operating in these studies occur during encoding and/or response, and the studies indicate that serial position effects occur in spatial as well as temporal short-term memory.

1.2.3 STM for motor sequences

The majority of work on motor memory has looked at single movements rather than at movement sequences. However a few studies have addressed short-term memory for sequences of movements. Typically these studies have used experimental apparatus consisting of a slide which moves frictionlessly along a horizontal track. Stops can be inserted at various points to limit the travel of the slide. The subject is blindfolded, and during the exposure phase moves the slide from a fixed starting point until it reaches the stop. This is repeated for a sequence of different movements. During the recall phase the stop is removed, and the subject attempts to reproduce the sequence of movements.

Several studies have found U-shaped serial position curves for errors in this paradigm. The first appears to be Burwitz (1974), and the effect has been confirmed by, for example, Wrisberg (1975) for reverse serial recall, Wilberg (1990) for free recall, and Magill and Dowell (1977) for serial recall. Magill and Dowell also found a clear effect of sequence length on recall for three and six item sequences, above which performance levelled out.

1.3 Motor Behaviour

Apart from the STM studies discussed above, much of the investigation of serial order mechanisms in movement has centred on timing. Rosenbaum, Kenney and Derr (1983)

studied timing and patterns of errors obtained from subjects executing sequences of finger taps, and present evidence that the control of rapid movement sequences can be hierarchical. Their stimulus sequences were arranged according to obvious hierarchical structures, with clear 'chunks' within the sequences. Both error and latency data show worse performance for the first item in a subgroup or chunk. Their theoretical claim is that sequences are stored in a hierarchical form and are unpacked at execution. A simple tree model of hierarchical storage predicts both the latency data and the types of error they found. The evidence does not support the hypothesis that rapid movement sequences are controlled by 'chained' associations from one movement to the next, either in planning or in execution.

Yu and Margoliash (1996) present empirical evidence that one form of serial behaviour is generated by hierarchically organised systems. They made multiple simultaneous intra-cell recording from zebra finch forebrains while the birds were singing. Zebra finch songs are organised hierarchically, with a clear distinction between individual notes and organisational units comprising several notes and known as syllables. One set of cells showed activation patterns uniquely associated with syllable activity, another showed precisely timed bursts uniquely associated with note identity. This indicates that separate cell assemblies represent the ongoing sequential activity at different levels of hierarchical organisation - in other words, the hierarchical structure which had been applied purely descriptively to zebra finch songs appears to be founded on actual hierarchical structure in the brain systems controlling song production.

1.4 Common Features of Serially Ordered Behaviour

The data reviewed above are striking in two respects. Firstly, the brain's strategies for dealing with serial structure appear to be rather more complex than would be suggested by the perhaps intuitive idea (the 'computer program' approach) of generating sequences by reading items one at a time from an ordered list or buffer. Secondly, similar patterns of data from experiments in very different modalities seem to indicate either that a common mechanism is involved, or that similar strategies have evolved to deal with the problem of generating and controlling serial order in different areas of behaviour. The purpose of this section is to abstract from the experimental data those

features which appear to relate specifically to serial behaviour and to be shared across several areas. The intention at this stage is to be purely descriptive. To proceed from the observation of commonalties in different behaviours, elicited under different conditions, to inferring shared or common mechanisms underlying them is to take a very large step. The implications of the observed commonalties are considered in the next section.

In all, looking at the data summarised in the previous section a total of nine common features may be discerned:

1. Co-activation of sequence elements. To Lashley (1951), one of the most striking implications of serial behaviour data was that the various responses making up a serial sequence (syllables or phonemes in a word, words in a sentence, etc.) are available and primed for production over a significant time period before and after the point where they should be produced. In speech production, the area which Lashley mainly focused on, many errors involve the movement of items from one position in a sequence to another. Items may be produced too soon or too late, implying that they are accessible both before and after their target position. Evidence for this type of 'window' of activation extending before and after the current position in a sequence of responses is also shown in spelling, typing and verbal STM. 'Co-articulation' effects in speech and typing offer additional support for this position.

Although items in sequences appear to be simultaneously active during sequence production, not all items are equally available - that is, items do not indiscriminately move from any part of the sequence (sentence, word or whatever) to any other part. Thus in verbal STM studies a common finding is that the farther apart two items are, the less likely an error on one is to involve the other.

2. Effects of serial position. Primacy and recency effects are found in studies of spelling, verbal and spatial STM and motoric STM. Brown (1991) points out that there is evidence from the TOT phenomenon that primacy and recency effects also occur in speech production. While many investigators have viewed the primacy and recency effects as evidence for the concurrent operation of separate memory systems, Greene (1986) concludes that the most likely explanation for the effects is the added salience of

end positions in an ordinal encoding scheme. The existence of both primacy and recency effects in domains which would probably be considered by most memory researchers to be largely distinct from verbal STM adds further force to the suggestion that the U-shaped serial error curve does not require the interaction of two or more separate subsystems. One would either need to postulate appropriate interacting subsystems within the systems subserving each of spelling, spatial STM, motoric STM and speech production, or one would have to provide evidence that primacy and recency effects arise from a unitary system in these cases but not in the case of verbal STM. Serial position effects will be discussed at some length in later chapters.

3. Effect of sequence and item length. In a number of areas a marked effect of the length of the sequence being recalled on the accuracy of recall is found. In spelling studies, the longer the word, the more likely an error. In verbal STM studies, the greater the *temporal* length of the target sequence, the greater the chance of an error.

4. Predominance of order errors. The data from speech errors, spelling, typing and short term memory all show a predominance of errors in the ordering of items over those affecting item identity.

5. Effects of repeats. Spelling, speech, verbal STM and typing data all show that repeats within sequences affect error incidence. Speech error data and verbal STM data show a specific detrimental effect of repeated items on order, rather than item identity, information.

6. Behaviour of double items. The (pathological) spelling and typing data both show the interesting effect of geminate (double letter) movement. This implies that close repeats involve a specific mechanism or 'schema', which is subject to errors in its positioning. Rumelhart and Norman (1982) have proposed that special purpose schemata are used to produce both doubled letters and alternated letters in typing.

7. Effect of similarity. Speech error and verbal STM data both show a detrimental effect of similarity between items on order information. Verbal STM studies show a clear effect of phonological similarity, which appears to lead to an increase in movement errors between similar items while leaving non-similar items unaffected. In

speech production, phonemic similarity again facilitates order errors, and similarity on other levels of description also has the same effect (syntactic class for words, syllabic position for phonemes) (Shattuck-Hufnagel, 1979).

8. Evidence for hierarchical control. The data for speech errors, short term memory and motor control all show evidence for the involvement of a hierarchical structure at some stage in the serial recall process. Shaffer (1976), in an interesting paper which addresses some of the similarities between serial processes in different types of behaviour, considers several putative models of serial behaviour and concludes that the evidence favours those which explicitly postulate hierarchical control.

9. Preservation of domain-specific constraints. Whenever constraints exist in a domain they are enforced when errors occur. This observation is most obviously true of speech production, where several different levels of constraints may be identified (e.g. phonotactic constraints, syntactic constraints at various levels) and where these constraints are enforced when errors occur on items at these levels. (i.e. words swap with words, NPs swap with NPs etc.). The preservation of consonant-vowel status in the spelling errors of GBD patients appears to be a similar effect, and there is an effect in the verbal STM literature which might also be seen as analogous in some respects: When 'chunks' exist in STM experiments errors occur which could be said to respect the chunked status of data, as for example when the first item in chunk A swaps with the first item in chunk B.

1.5 Implications

The occurrence of very similar features in the data from several different areas of psychological investigation is interesting. However, care is required in interpreting these correspondences. Some possible explanations for similar phenomena occurring in different psychological areas might be:

1. A single mechanism, or set of mechanisms, is shared by all the psychological competencies involved. For example, commonalties between language functions and verbal STM might indicate that some language-based system employed by both was having a large effect on error production. The inclusion of motor STM would make

it less likely that specifically linguistic functions are shared, but might indicate that it is a shared low-level motor control function which causes the errors.

2. A few separate mechanisms might be involved, with all the competencies discussed above using one or the other of them. The obvious question in this case is how it is that these separate mechanisms all produce such similar phenomena.

3. The critical errors produced in each of the competencies may be the result of totally separate mechanisms. The question posed above is even more pertinent in this case, as the number of mechanisms which putatively operate in the same manner is larger.

The commonalties between separate mechanisms in possibilities 2 and 3 might however be explained by similarities in their task, or by their evolutionary background (They may for example all have evolved from the same type of underlying mechanism).

Both the shared subsystem (possibility 1) and distinct subsystem (possibilities 2 and 3) options are equally interesting. Identifying a shared subsystem would have implications for theories of all the psychological competencies involved, while the suggestion that a number of substantially separate brain systems all operate in a very similar manner would certainly prompt a number of questions about how this could come to be.

It would be difficult to adequately distinguish the above possibilities using traditional psychological experiments. This is an area where the construction of models will be particularly useful. If the common features identified above are a heterogeneous set which require an ad-hoc set of architectural features to explain them, the unlikeness of such a coincidence of features in several distinct processing systems would strongly suggest that a set of common subsystems were responsible. If, on the other hand, the common features of the data are explained, en masse, under a single processing model, the proposition that such a processing scheme might arise independently in more than one separate subsystem is more reasonable. The next chapter will examine some possible schemes for the generation of serial behaviour with this question in mind.

2

Computational models of serial order generation

The formalisation of a theoretical model of cognitive function as a computer program is an important research tool in cognitive science, and serves several functions. As well as providing an unambiguous formal specification, it demonstrates that the model is sound and does what it is intended to do, which is not always obvious if anything other than the simplest interactions are involved. A further use is as an aid to theorising; it is often easy to take for granted seemingly trivial aspects of cognitive processes, and the discipline of formal specification and programming help ensure a rigorous approach. There is if anything even more need for care in modelling serial processes, as the ease with which serial behaviour may typically be produced within computational frameworks may itself be misleading - it is unlikely that the brain has recourse to anything resembling the serial primitives of computer programming languages, which may trivialise the difficulties involved.

There have been numerous attempts at modelling serial behaviour in various psychological domains. These domains are often rather small, as cognitive models tend to address particular limited aspects of behaviour rather than general serial processing. Nonetheless, some basic common principles for the generation of serial order emerge.

This chapter reviews some of the major serial models, computational and otherwise, with a view to elucidating these principles. The first part of the chapter reviews five important approaches to serial ordering: Structural, Chained, Time-delay, Distinctiveness and Competitive Queuing models. A number of common threads which occur in different models are identified and discussed in the light of the common features in the psychological data reviewed in the last chapter. The second part of the chapter concentrates on a more detailed discussion of the last of these approaches, Competitive Queuing, which will form the basis for the modelling work of subsequent chapters.

2.1 Structural Models

This first section concerns models which represent serial order by imposing a structure on stored information.

2.1.1 Serial Buffer Models

The standard structural representation of serial order assumes a series of 'boxes' or 'slots', each of which can hold one response and which are accessed one at a time to produce the response series. This idea has been encouraged by the computer metaphor for complex or intelligent systems - it is the most obvious way to generate serial behaviour using a standard computer language. In fact considerable extra work is usually required in order to produce serial behaviour in any other way on a computer.

Conrad (1965) proposed a model of short-term verbal memory as a series of 'boxes' or 'slots', into each of which a subject enters an item in a presented sequence. To recall the sequence the subject simply reads out the contents of each box in order. Conrad's main purpose in proposing such a simplistic model was to make the point that order errors in verbal STM could be artefacts created by acoustic confusions between similar sounding words, and hence could be explained by a model which had no mechanism for intrinsic order. This argument has been challenged by evidence that order errors occur during storage rather than during retrieval (e.g. Murdock and vom Saal, 1967), and by the fact that recall for order information and for item information shows different profiles of error behaviour (Fuchs 1969, Healy 1974, Aaronson, 1968) and can be differentially manipulated by varying experimental conditions (Wickelgren, 1965, Bjork and Healy, 1974).

In addition, this model cannot can not explain the types of error produced when memory breaks down. The basic prediction of such a model is that, on breakdown, item identities may become degraded. While this would presumably lead to errors, the serial buffer specification on its own can say nothing about what form such errors would take. The buffer model needs to be augmented with additional mechanisms handling structural and algorithmic information to explain the structure of the errors. Conrad, for example, does this implicitly by making the assumption that the same mechanisms that

lead to acoustic confusion in speech perception also lead to erroneous recall of items from boxes when those items are degraded by decay, and there is thus considerably more to his model than an ordered set of boxes.

2.1.2 'Wickelphones'

Wickelgren has proposed a series of models of speech production in which order is generated implicitly among a set of simultaneously activated items (generally phonemes) by a constraint system. The representations for items in these schemes are context dependent. For example, if the word *cat* comprises the ordered set of three phonemes /k/ /A/ /t/, the representational units for these phonemes would be $_*k_A$, $_kA_t$ and $_At_*$ respectively, where $_kA_t$ means the phoneme /A/ preceded by /k/ and followed by /t/, and * indicates a word boundary. Each phoneme has a different representation depending on the phonemes which precede and follow it, so a very large number of representational items are required by such schemes. (Such representational items are often termed 'Wickelphones').

As well as suffering from the general problem of 'structural' models - the lack of an explicit algorithm for converting the stored representation into actual serial behaviour, and thus the inability to make any concrete predictions - the main problem of 'Wickelphone' models is the enormous proliferation of distinct elements required to implement such schemes. Even in the rather limited domain of case-free letters, where only 27 output responses exist (26 letters and 'space'), 19683 (27^3) separate representational units are required if only one item on each side of the 'target' is used. With two letters of context in either direction the number rises to over 14 million. Clearly if other domains are included the number of representational units will soon become even larger - consider, for example, the number of representations required to allow all possible groups of five English words. Each of these representations must be separately learned and stored, a task which in itself requires some effort to explain. However, the most serious problem with this scheme is that, like Conrad's 'boxes', it does not explain the common types of errors on serial tasks, such as exchanges, insertions and deletions.

2.2 Chained Models

The behaviourist approach of seeking the explanation for behaviour in terms of stimulus-response pairs naturally suggests a mechanism for serial behaviour - each response in a sequence should be the stimulus for the next. A number of current models have this character.

2.2.1 TODAM

A number of models of non-serial memory have used operations on vectors, typically addition, convolution and/or correlation, to store and retrieve information (e.g. Eich, 1982, Murdock, 1982). Murdock's "Theory Of Distributed Associative Memory" (TODAM, Murdock 1982) is one such model. Items to be stored in memory are represented by vectors with random elements, so that each item is represented by a distinct vector. Items are stored in a memory trace by vector addition, and associated pairs of items are stored by adding the convolution of their vectors to the memory trace. Recall of one of the pair is performed by correlating the trace with the vector representing the other. The basic TODAM model does not allow serial storage and recall. The model is augmented by Murdock (1983) and Lewandowsky and Murdock (1989) to deal with serial recall tasks using associative chaining. A sequence of items is stored by adding to the memory trace both the vector representing each item and the association (represented as the convolution of vectors) between each item and its successor. As each item is retrieved during sequence production, it acts as the cue for its successor. An initial 'start' signal is needed to act as cue for the first item in the sequence and start the whole process off.

The most common criticism of models which use chaining is that they are unable to continue with correct recall following an error. This is in contrast to normal human recall which often proceeds normally following an error in mid sequence. The serial version of TODAM is able to circumvent this difficulty by using a two-stage process for recall. The recalled item is only an approximation to the original stimulus because of intrinsic inaccuracies and because random noise is added during the retrieval process. As a result, the output from the cued recall operation needs to be 'deblurred': A similarity metric determines the actual item representation closest to the approximate

output and this is used as the response. If no item is sufficiently determined by the similarity metric recall is assumed to have failed. When recall is successful the 'deblurred' item representation is used as the cue for the next item's recall, but when recall fails the approximation yielded by the first stage recall operation is used as the cue. Thus a cue is always available, although it may only be an approximation to the ideal cue.

The model produces primacy and recency effects similar in shape and magnitude to those observed in human serial recall (Lewandowsky and Murdock, 1989). The mechanism for recency is the mechanism adopted in the model for general forgetting - as each new item is added to the memory vector, the previous state of the vector is reduced in magnitude by a constant amount. This has the effect of lessening the contributions of items further in the past compared with that of more recent items. The primacy effect on the other hand is due to a mechanism which is entirely ad-hoc: The weighting parameters governing the influence of each new item on the memory vector are reduced in each successive serial position.

Although the model is able to continue recall following an error there is no obvious mechanism whereby ordering, and especially exchange, errors might occur. The lack of an explanation for such a ubiquitous phenomenon in human serial behaviour is a serious problem. Murdock (1995) addresses it by introducing remote associations to produce a compound chaining model which does produce exchange errors. Problems remain with the TODAM approach, however. No mechanism is available to explain the effects of phonological similarity in short-term memory, for example. Further, the model is prey to a number of other criticisms of associative chaining. The use of chaining in short-term memory models faces difficulties in addressing data which shows associations between serial positions and the items occupying them (Henson, submitted). Henson, Norris, Page and Baddeley (1996) also point out that whatever the amount of past state retained, a chaining model will always suffer from the problem that an error will result in the following item in the series being presented with an incorrect cue. This must result in higher error probability following an error, but Henson et al. find a level not significantly different from chance.

2.2.2 Recurrent Connectionist Network Models

Connectionist (or neural network) models employ networks of units ('nodes') which each perform simple computations, such as summing their inputs and applying thresholds. The units are interconnected by links which are capable only of transmitting a single scalar quantity (the 'activation' level of the transmitting unit), and have a single parameter, their 'weight', which is a scaling factor applied to the activation level being transmitted. Units are assigned activation levels as a function of the weighted sum of inputs they receive from other units, and in turn may send output signals dependant on their activation level along weighted connections to further units. Since the units themselves are unable to perform complex computations, it is the arrangement of connections between them which does most of the work in transforming input to output. Processing in such models is thus controlled by the combination of the topographical layout of connections and their scaling weights. Often models include learning procedures which enable the connection weights to be modified in order that the network can learn to process data. The simplest such rule (due to Hebb, 1949) increases the weight on the connection between two units to the extent that their activation levels co-vary. More complex learning procedures such as error backpropagation (Rumelhart, Hinton and Williams, 1986) allow learning in networks containing units which are not directly connected to either input or output ('hidden units'). A full introduction to connectionist modelling is beyond the scope of this thesis (see e.g. McClelland and Rumelhart, 1986, Bechtel and Abrahamsen, 1991); the necessary concepts will be introduced as required.

Standard neural networks are very effective at tasks which involve mapping from an input pattern to an output pattern. One obvious way to achieve sequential behaviour from a network is thus to cause the network to map from its current internal pattern of activation (state) to a new state. Many states may then be chained together using this mechanism to form a temporal sequence. This idea forms the basis of the recurrent networks developed by Jordan (1986) and Elman (1990).

Jordan's (1986) architecture consists of a standard three-layer feed-forward network, with input, hidden and output layers of units, and connections between them trained

using the backpropagation algorithm. The input layer nodes are split into two groups. One group represents the 'plan', a pattern which is held constant throughout the sequence and specifies which particular sequence the network is to learn or produce. The other group of input nodes specifies the current state of the network, and these nodes are activated by recurrent connections from the state nodes themselves and from the output nodes (see Figure 2.1 a). This allows the activation pattern of the state nodes to depend on the previous output from the network and on the previous state. Elman's (1990) architecture differs in that, rather than feeding back the activation pattern of the output layer, the Elman net feeds back the previous activation pattern on the hidden units. This is achieved by copying the activation of the hidden units to a set of 'context' nodes at each time step (Figure 2.1 b). Since the current state of systems such as these depends to a certain extent on previous states (and thus on previous input) their chaining is not necessarily based just on the current state but may be a function of the entire input and state sequence to date.
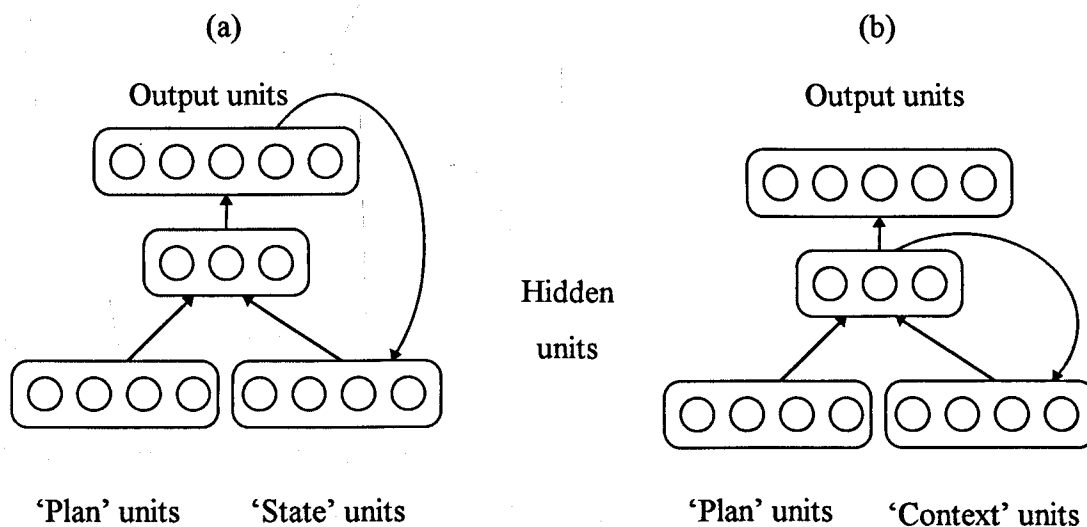
Figure 2.1. *The recurrent network architectures of Jordan (1986) (a) and Elman (1990) (b).*

Similar recurrent architectures have been developed within other neural network paradigms. Amit, Sagi & Usher (1990) have produced a model of the Sternberg paradigm (Sternberg, 1966) using a modified Hopfield net. Hopfield nets can be made to settle into stable 'attractor' states, where the pattern of activation across the network

becomes self-supporting and does not change. However this particular model includes connections which map from one attractor state to another. When the network begins to settle into one attractor these connections tend to force it into the next. This allows a chain of attractor states to be assembled, each one providing the stimulus for the next. Amit et al. use this ability to model the rehearsal of sequential input.

As chaining models, Jordan and Elman networks suffer from the same disadvantages as the serial version of TODAM. Elman nets have the advantage that if the mechanisms leading to errors are located at the output of the system and the chaining cues originate more centrally, sequencing may be able to continue following an error. Errors may occur at the output due, for example, to damage to or excessive noise on the weights from hidden to output layer, to damage to the output layer itself, or to disruption of the output layer by noise or by outside influences. Despite such errors, the *cue* for each item remains undamaged at higher levels in the net. Other factors such as the difficulty of adequately explaining transposition and exchange errors and the lack of any association between serial position and content still argue against chaining in this form, as in others.

2.3 Time-delay Models

2.3.1 Lee and Estes' Perturbation Model

Lee and Estes explicitly introduce the passage of time as the organising principle in serial behaviour, and over a number of years they have developed an influential model of serial verbal memory based on the idea of temporal perturbations in reverberatory cycles. The initial version of the model (Estes, 1972) rejects the idea of direct item-to-item chaining, and instead proposes the idea of a 'control element', which represents a sequence in its entirety. Each element in the sequence is activated by association with the control element rather than with other sequence elements. The logical structure of a representation of the sequence '123' in memory is thus as shown in Figure 2.2.

C

1   2   3
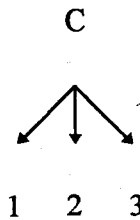
Figure 2.2. *The representation of the sequence '123' in short-term memory. 'C' is the control element for the sequence.*

Estes suggests that the sequence is maintained in short-term memory by a reverberatory loop mechanism: After presentation, and while the control element for the sequence remains active, the representation of each item is periodically re-activated at a constant rate, so that the sequence is endlessly repeated. With a certain probability, each item may be perturbed in its timing so as to occur either sooner or later than it did on the previous cycle, so that over time the position of an item in the sequence may drift. The start and end points of the sequence are fixed so that the first and last items may each only be perturbed in one direction.

This reverberatory rehearsal of the sequence is sufficient to keep it present in short-term memory. However, maintaining the sequence in this way is an active process and is thus not well suited to long-term storage. If the sequence is to be stored in long-term memory and later recalled a different representation is required. Estes proposes that after the reverberatory loops maintaining STM for the sequence are set up, inhibitory connections start to develop from early items to later ones (Figure 2.3), establishing long-term memory for the sequence. Recall from LTM involves activation of the control element C, which activates the individual items, 1, 2 and 3. The inter-item inhibitory connections establish a gradient of activation levels over the items such that 1 is most active, 2 somewhat less active (it is inhibited by 1) and 3 relatively inactive (it is inhibited by both 1 and 2). This gradient initiates the correct sequencing of the items: The first to be recalled is 1, as it is fully active. Following its activation item 1 is assumed to be removed from the 'competition' for recall, and the influence of 1 on 2 is removed so 2 becomes fully active. 2 is thus the next item to be recalled. The influence of 2 on 3 is then removed and 3 is recalled. Following this initial run through the

sequence the reverberatory mechanism maintains the sequence in STM as long as is required.

Some of the explicit reference to reverberatory loops is replaced in the later version of the model (Lee and Estes, 1977, 1981) by an encoding scheme which is similar in nature, but rather more formal. Each item is represented by an 'entry' in memory detailing its identity and its position. The positional coding is still time-based and probabilistic, and items can shift in time as in the original version.
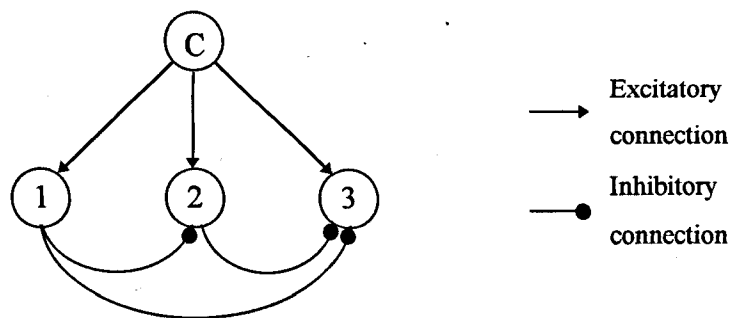
Figure 2.3. *The inhibitory chaining scheme suggested by Estes for sequence initiation from long-term memory. Earlier items inhibit later ones until they have been produced.*

The main problem with this model is the lack of any specified mechanism to implement the reverberatory re-activation of items. (The inhibitory chaining mechanism assumed for LTM recall is itself just as inadequate as other response chaining models of sequence recall, as discussed elsewhere. This particular mechanism is discussed in more detail below in connection with Rumelhart and Norman's typing model). The fact that read-out from LTM and immediate recall require different mechanisms is also somewhat unsatisfying. There does not appear to be a mechanism for material to intrude from outside STM. (The only obvious mechanism would be subjects guessing). Finally, the model offers no mechanism for phonological confusions. This would presumably have to be added to the model in the form of an output phonological mechanism.

2.4 Distinctiveness Models

2.4.1 Johnson's Distinctiveness Model

Johnson (1991) proposes a model of serial learning based on associations between each item in a sequence and a code representing the ordinal position of that item in the sequence. Johnson's model is a concrete realisation of a proposal originally made by Murdock (1960) that the ease of recall of any item in a learned sequence is dependant largely on its distinctiveness, which in turn is dependant on its position in the sequence. Murdock proposed a quantitative index of the distinctiveness of one item in a set. Johnson (1991) makes Murdock's index the basis for a concrete model. Each item is assumed to be associated with every ordinal position in the sequence, but with greater strength of association for the position in which it should be recalled than for distant positions. Johnson makes some straightforward assumptions relating the probability of recalling an item in a particular position to the strength of associations between that position and the item, and is then able to show a remarkably good fit to data in several paradigms, including the usual serial position curves for different list lengths and different degrees of learning (modelled by altering the mean association strength) and the effect of number of repeated items on overall performance.

Although the model is defined in concrete terms, its major disadvantage is the absence of an explicit recall algorithm by which the pattern of associative strengths may be converted into actual serial recall. There are thus many aspects of serial recall about which the model is silent. For example, nothing can be said about the way in which different types of error (for example, deletions, transpositions or insertions) can result from the associative structures defined by the model. The model is also unable to say anything about errors involving items not present in the stimulus list (item errors), and makes incorrect predictions regarding the relative incidence of anticipatory and perseveratory errors in serial recall (subjects make very few perseveratory errors, and almost never repeat an item without a large intervening distance; the model predicts symmetrical occurrence of perseveratory and anticipatory errors, and does nothing to prevent the close or even immediate repeat of items). However, Johnson's representational scheme has much in common with several other models which are

discussed in section 2.7, and these models may be viewed as implementations of ideas similar to Johnson's.

2.5 Competitive Queuing

A class of models exists which separate 'item', (sequence membership) information from order information, sequence membership being denoted by the activation of item representations, and the order of these items by their relative activation levels. These models differ in the mechanism which generates the gradient of activations across items. However, all share the characteristic that several items are active in parallel during sequence production, that the ordering of these items is determined by their relative activation levels, and sequencing is achieved by repeatedly selecting the most active item for output and then removing it from the pool of competing items. Houghton (1990) uses the term 'competitive queuing' (CQ) for such systems. Houghton's CQ architecture has several specific features, but the term is a useful one to describe the broader class of models which generate sequential behaviour in this way.

CQ models comprise three main elements, shown conceptually in Figure 2.4:

1. A set of **refractory item representations.** These constitute a pool of available individual items from which the sequences generated by the system are composed. It is central to the CQ approach that these items are refractory - that is, when an item is produced as part of a sequence, it becomes temporarily unavailable for further use.

2. An **activating mechanism**, which activates those items in the pool of available items which are required for the particular sequence being generated. The relative activation levels of these items determines the order in which they will be produced, so the activating mechanism is important to the sequencing process. Depending on the type of model, the relative item activation levels generated by this mechanism may be static throughout the sequence generation process or they may change over time.

3. An **attentional output mechanism**, which, whenever a new item is to be produced, selects the most active of the pool of available items and passes it to the output of the system.
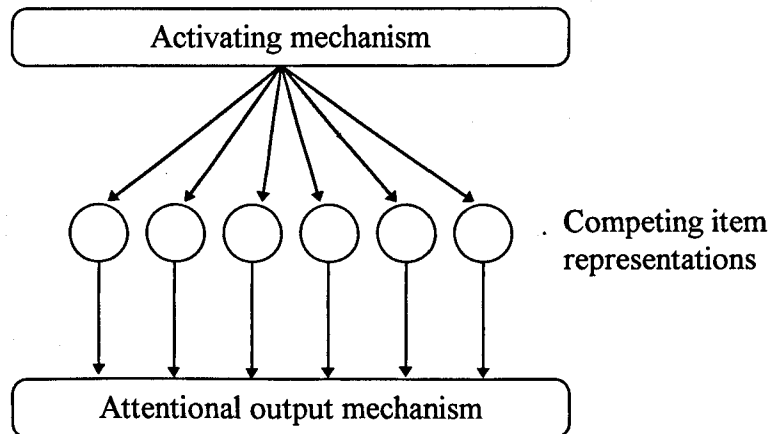
Figure 2.4. *The basic elements of Competitive Queuing models.*

The process of generating a sequence in such a model involves the activating mechanism generating a *gradient of activations* over some subset of the item representations, such that the item to be produced first is the most active and the item to be produced last is the least active. The output mechanism then repeatedly selects for output the most active item. As each item is output it becomes refractory, and hence temporarily unavailable. In this way the activated items are output in order of their relative activation values, from the most to the least active. All CQ models comprise these basic elements and operate in essentially the same manner, although the implementational details vary from model to model and additional components may be added.

The CQ paradigm has a number of attractions as a basis for models of serial behaviour in the light of the data of Chapter 1: It does not involve associative chaining, it naturally gives rise to the usual effect of sequence length - longer sequences are more prone to error - and perhaps most importantly it gives a clear and simple account for the high incidence of ordering errors in general and exchange errors in particular in serial tasks, which follow from the fact that response chaining is not used, from the activation-based

selection approach which transforms any uncertainty in the relative activation levels of items into movement within the sequence, and from the refractoriness of items which prevents repetition errors and encourages exchanges. Finally, simple versions of the model have given a good account for the higher incidence of errors in medial positions of sequences. The mechanisms behind all these effects will be discussed in detail below.

Not surprisingly CQ models have been successful in a range of psychological domains, the general principles of the approach often being arrived at independently by different authors. Models have covered the areas of verbal short-term memory (Milner, 1961, Burgess and Hitch, 1992, 1996, Glasspool, 1995, Page and Norris, 1997, submitted, Henson, in press a., Brown, Preece and Hulme, submitted), typing (Rumelhart and Norman, 1982), speech production (Dell, 1986, 1988, Houghton, 1990, Hartley and Houghton, 1996, Vousden, 1996), spelling (Houghton, Glasspool and Shallice, 1994, Glasspool, Houghton and Shallice, 1995, Shallice, Glasspool and Houghton, 1995) and general sequential behaviour (Grossberg, 1978, Norman and Shallice, 1980, 1986, Cooper, Shallice and Farringdon, 1995, Cooper and Shallice, 1997; in submission).

2.6 Interim summary

This completes the review of the important approaches to the generation of serial behaviour in computational models. A number of common threads can be identified running through the models discussed above, which this section picks out in order to provide a complimentary perspective on the different approaches. The following common features are particularly evident in several of the approaches discussed above:

Chaining

A number of models take the view that seriality must be generated by making the current response dependent on the previous response. Two approaches are possible, which might be termed 'excitatory' and 'inhibitory' chaining. Excitatory chaining occurs when one item cues the production of the next. The element which cues production of the successor item may be the actual output, in response-chaining models such as Jordan's, or it may be some internal representation as in Elman's model or the serial version of TODAM. The former is strongly argued against by the fact that

sequence production often proceeds normally following the occurrence of an isolated error. The latter form may be more resistant to catastrophic failure following a single error (depending on the locus of the error), but remains prey to the other problems of chaining accounts.

The question of whether sequencing is carried out by chains of associations has been vigorously debated since Lashley's (1950) seminal paper on the problem of serial order. However, a great deal of evidence has now amassed which strongly suggests that very little, if any, chaining of item to item occurs in serial behaviour. Data from all of the modalities discussed in Chapter 1 show evidence that sequencing is not achieved by straightforward chaining of associations, the most obvious being the occurrence of exchange errors occur. Such errors are of the form ABCD → ACBD, where B and C exchange places. A simple response chaining account would predict that when the erroneous C is produced in the second position, the next response to be activated would be D, via the C to D association. Simple chaining theories would therefore predict that exchange errors should be rare. However, such errors are very common in slips of the tongue, slips of the pen, GBD spelling, typing and verbal STM. Evidence against response chaining also comes from error and timing information in tapping tasks (Rosenbaum et al., 1983). A further difficulty is the problem of explaining the speed at which human subjects can produce sequential behaviour. Thus for example chaining from completion of one action to initiation of the next cannot explain the speed at which serial behaviour can proceed in skilled typing.

Young (1961; 1962) presents evidence that the stimulus for the production of a response in a sequence in verbal STM is not the preceding response but the ordinal position of the response in the sequence (the so-called 'ordinal position hypothesis'). Ebenholtz (1963) supports this position by demonstrating that inter-item associations are not necessary for serial recall. Young (1968) finds evidence in long-term memory experiments which supports the ordinal position hypothesis for LTM too. Learning paired associates (for example, the pairs EF, CD, AB) does not facilitate the later learning of a list which, if item-to-item chaining were being used, would require the same associations (eg. ABCDEF). Young concludes that associations are to 'place markers' which indicate ordinal position.

Slamecka (1964), in a review of the experimental literature, concludes that serial learning proceeds by first learning the identities of the items involved and then fixing their sequential positions. The associations which are learned in serial tasks are most likely to be between each item and "some distinctive symbol designating its relative position in the list", rather than between consecutive items. More recently, Henson, Norris, Page and Baddeley (1996) present an argument against chaining and in favour of models using positional cues. They find that the probability of an error in serial verbal STM recall depends only on the position of the item in the sequence, and not on whether or not an error occurred in the previous position. Any theory which holds that any part of the stimulus for a response is the production of the previous response must predict that an error is more likely if the previous response is erroneous, since the stimulus will then be partially degraded. Additionally, evidence from 'protrusion' errors in verbal STM (Henson, submitted) suggests that an association exists between each serial position in a sequence and the item to be produced in that position, which is difficult to reconcile with a chaining approach even when chaining is performed on internal representations rather than outputs. Despite the tenacious hold of the item-to-item chaining model on the imagination of psychologists, then, the evidence appears to rule out any form of response chaining mechanism as a real contender.

Control elements

A number of models use a specific element to represent an entire sequence. This approach is adopted by Jordan and Elman in chaining models and by Estes (1972) in the active (short-term) part of his reverberatory model. The 'activating mechanism' of CQ models also constitutes a control element of this type since it is responsible for activating all and only the items within the sequence and thus represents the sequence in its entirety. If the separate elements of a sequence can all be tied to one common control element the problem of storing different sequences containing the same items may be more easily addressed. Additionally, the evidence reviewed in Chapter 1 suggests that at any point during the production of a sequence a number, if not all, of the component items of the sequence are simultaneously active, or in some sense available for output. Thus errors often involve an item being generated before or after its target position in the sequence, and in some cases (in spelling errors, for example)

items which should be produced at one end of the sequence may even turn up at the other end. The phenomenon of co-articulation also suggests that items are available for production before they are actually produced. The implication of the parallel activation of a sequence's constituent items is consistent with the 'control element' approach.

An advantage of the control element approach is that it allows the possibility of hierarchical organisation within sequences, identified as a common feature in Chapter 1. An entire sequence may be represented by a single control element. Nigrin (1993), Page (1993) and others have used this approach in producing hierarchical sequence generating models based on Grossberg's ART architecture, and a similar feature is evident in the hierarchical structure of the Contention Scheduling model. The possibilities have yet to be fully explored within the CQ framework, however.

'Select-Inhibit' sequencing

If a 'control element' rather than associative chaining is responsible for activating the component items of a sequence, some other means is required to produce them in the correct order. CQ models use a form of sequencing in which the most active of a set of competing items is chosen for output and is then inhibited. This mode of sequencing is also implicit in models which use inhibitory chaining, for example Estes (1972). The great advantage of this form of sequencing from the point of view of modelling human data is the good account it can give of ordering errors, generally the most prolific error class. Without the post-production inhibition of items errors are likely to result in repeats, which are not a common feature of human performance. The inhibition of produced items also explains the facilitation of errors by repeats, and has been used to explain the odd behaviour of doubled items in sequences (Houghton, Glasspool and Shallice, 1994, see also Chapter 4).

2.6.1 Relative merits

Serial buffer models are at best only metaphorical, with little explanatory power. Chaining and timing models suffer from a number of disadvantages which can only be partially overcome with some ingenuity. There is very little evidence that any form of chaining, in particular, takes place anywhere in behaviour. The approach which appears

to hold the most promise for explaining the general features of serial behaviour identified in Chapter 1 is CQ, which defines both a representational scheme and an algorithm for generating sequential output from it. The approach does not require the use of chaining to drive sequencing, and it incorporates the benefits of a higher level control element and 'select-inhibit' sequencing. CQ thus adopts what appear to be the most useful features of the other approaches reviewed while avoiding potential pitfalls. Moreover, it is able to provide an account for the 'core' common features of serial behaviour identified in Chapter 1 - the effect of sequence length, the predominance of ordering errors, the existence of exchange errors, co-activation, and 'primacy' and 'recency' effects - within a single consistent framework. Given the prevalence of these effects and the difficulty which some other approaches have had in accounting for them this is an important consideration. As a 'performance' model, with a well defined algorithm as well as a representational scheme for serial information, the approach also has the benefit of producing detailed quantitative results for direct comparison with empirical data.

Given the suitability of CQ for modelling human serial behaviour, subsequent chapters will concern models of the CQ type. There is however considerable variation in the implementational details of CQ models. The remainder of this chapter examines the paradigm more closely and discusses several CQ models in detail, in order to elucidate the critical differences between implementations and their impact on the operation of the models.

2.7 Competitive Queuing: A detailed analysis

The most important differences between different CQ models concern the nature of the activating mechanism (see Figure 2.4). There are two main variants:

1. **Static** mechanisms. These set up a single activation gradient over items which is then held static during the course of sequence generation. Sequence production is accomplished by virtue of the refractoriness of item representations - as each item is produced it becomes refractory and is removed from the competition for output in the next sequence position, where the next most active item will 'win'. In such a

system it is desirable that items should remain refractory for a long period, as items produced near the beginning of the sequence will have a large input from the activating mechanism which will still be present towards the end of the sequence, when only the fact that the item is still refractory will prevent it re-appearing in the sequence. The use of a static activation mechanism places two constraints on a CQ system which may or may not be disadvantageous depending on the details of the model: Firstly, the activation level of each successive item in the sequence must be lower than its predecessor. Secondly, unless there is some means to control the refractory periods of individual item representations (which is generally assumed not to be the case), it is not possible to cater for sequences which contain repeated items.

2. **Dynamic** mechanisms. These allow the activation gradient set up over the item representations to vary during the course of sequence generation. The fundamental effect of this is to reduce the reliance on the refractoriness of items for correct sequencing behaviour - some of the work of sequencing is done by the activation mechanism in varying the relative activation levels of items. This can render the sequencing process both more reliable and more flexible. A number of factors may make sequencing more reliable: The number of items competing for output in each sequence position may be reduced by arranging that only a few items are strongly active in each sequence position; the activation level of each successive 'winning' item may be brought up to a similar high level, improving noise immunity; and input from the activation mechanism to items early in a sequence may be reduced as sequencing progresses, reducing the need for an excessive refractory period to prevent perseverative errors. Depending on the details of the model, it may be possible to make sequencing more flexible by allowing repeated items to be produced in the same sequence. This requires items which have been produced (and are therefore refractory) to be reactivated by a large activating input.

Section 2.6 discussed the 'ordinal position hypothesis' (Young, 1961, 1962) as a contrast to the idea that serial behaviour involves associative chaining between items in a sequence. The hypothesis is that the stimulus for the production of a response in a sequence is some representation of its ordinal position in the sequence. Henson and

Burgess (1997) point out that two distinct versions of this hypothesis may be distinguished: The position of an item in a sequence may be represented in a *relative* manner, in terms of its position relative to the other items in the sequence (Henson and Burgess term this *ordinal* representation) or in an *absolute* manner, in terms of its absolute position in the sequence regardless of other items (Henson and Burgess's *positional* representation). Applied to the activating mechanisms of CQ models, the static form may be classed as encoding an ordinal representation of sequence position, since the position of an item in a sequence depends only on its activation level relative to the activations of the other items. (Thus the third most active item will be the third item in the sequence.) The dynamic form of activating system introduces a positional element to the representation of ordering in the sequence - since the signal generated by the activating system varies with time, the state of the signal at any moment encodes the current position in the sequence, irrespective of the states of the item representations. However, an ordinal component remains. Although the gradient of item activations varies, it is still the local gradient at any moment which defines which item will be produced in that position. The rate of change of the activation gradient is thus of interest. If the gradient changes very slowly, the system is not very different to a static system and the ordinal component of order representation may be said to dominate. If the gradient changes more rapidly it becomes easier to discriminate successive positions in the sequence from the local state of the gradient, and the positional component comes to dominate. In the extreme case only the target item would be activated in each sequence position, and a fully positional representation system would result. In models which use a dynamic activating mechanism a smooth continuum of possibilities exists between the two extremes of a fully positional or fully ordinal representation of item position. This may be thought of in terms of the distinctiveness of consecutive states of the mechanism - the more distinctive, the more rapidly the activation gradient is changing and the further towards the fully positional end of the continuum the model may be placed.

In the dynamic type of model, the activating mechanism can be viewed as the combination of an *input signal* embodying information about the position in the sequence and an *activation function* which determines how this information affects the

activation levels of items. A similar distinction is used in connectionist frameworks, where the activation level of a connectionist node is determined by the net input to the node transformed by an activation function. Houghton (1994a) considers the ways in which the distinctiveness of consecutive positions may be increased in CQ models, and points out that two routes are available, corresponding to either increasing the discriminatory power (or positional resolution) of the activation function, so that consecutive states of the same input signal may be more easily discriminated, or increasing the distinctiveness of consecutive states in the signal itself, allowing the same activation function to more easily distinguish consecutive states. Improving positional discrimination can improve both the flexibility and the robustness of the sequencing mechanism, by an extension of the arguments mentioned in the discussion of dynamic forms of activation mechanisms, so this aim has been pursued in a number of CQ models.

The remainder of this section briefly discusses the most important computational models which use the CQ approach, classifying them according to their use of static or dynamic activating systems and, if dynamic, by the degree of discrimination available between consecutive positions - that is, the degree to which the positional component of the dynamic activating mechanism dominates the ordinal component.

2.7.1 Static models

The models in this section use a static form of activating mechanism. Their representation of sequence order is thus ordinal.

Rumelhart and Norman's typing model

Rumelhart and Norman (1982) describe a model of typing which takes a CQ approach to sequence generation. The model aims to simulate the motor control of a skilled typist in some detail, and includes an elaborate model of the control of each finger during typing. However, this motor response system is driven by a sequencing mechanism which generates the overall sequence of letters to be typed, and it is this mechanism which is of interest in the current context. The model contains a set of nodes representing letters (one node per letter) which form the input to the response

generation mechanism. Words are assumed to be represented within the model as 'schemas'. When a particular word is to be typed, its schema is activated, resulting in a set of connections being established to and amongst the letter nodes required to type the word. The general scheme is the same as Estes' approach to retrieving sequences from long-term memory in the reverberatory model discussed earlier: All the letters of the word are activated in parallel, but inhibitory connections are established such that each letter inhibits all later letters. The earlier a letter occurs in the word, the fewer prior letters it has and the less inhibited it is. The gradient of letter activations is fed directly into the motor control response system, where it generates finger movements towards letter keys, the degree of movement towards each key depending on the activation level of the corresponding letter node. As the gradient of activations in this model is quite sharp, most of the movement is directed towards the current target key at each step in typing the word, although the up-coming letters also influence movement. When a finger is within a threshold distance of a key it is struck, and the corresponding letter looses activation input from the word schema.

The model makes the same types of errors at the letter level as skilled human typists (see section 1.1.2) when a small amount of random noise is added to activations in the model to cause uncertainty in the response triggering mechanism. Transposition errors, common in typing, are produced by the model due to the combination of pre-activation of up-coming letters (which provides the opportunity for anticipatory error) and the effective inhibition of letters following their production (leading to typically movement, rather than perseverative, errors). Much of the effort of modelling is directed at the phenomena resulting from co-activation of up-coming letters along with the current target letter, and the good performance of the model in this area reflects the CQ approach of generating a sequence of responses from the parallel activation of a set of response items - finger movements in the model smoothly accommodate the multiple concurrent goals to produce efficient trajectories in the same way as a human typist. However, the use of a static activation gradient means that the model cannot easily produce repeated letters. Words containing repeated letters have to be treated as a series of sub-word 'chunks', each containing no repeats.

An interesting aspect of the model's operation is the mechanism used for double letters. Section 1.1.2 introduced the interesting class of errors which typists make on these - the attribute of 'doubling' appears to be detachable from the particular double letter, so that for example the wrong letter is doubled. Rumelhart and Norman interpret this as implying that double letters are represented as a single occurrence of the letter with a special purpose 'doubling schema' bound to it. It is assumed that the schema may become bound to the wrong letter under the influence of noise. This leads to the correct modelling of doubling errors, but is an ad-hoc addition to the model. (This mechanism will be discussed in greater depth in Chapter 4).

The primacy model of verbal STM

Page and Norris's (1997, submitted) Primacy model of verbal short-term memory uses the basic static CQ approach of repeatedly selecting the most active of a set of items and then inhibiting it. In this case the items are words, and the model addresses the immediate recall of a sequence of words in verbal short-term memory tasks. Apart from the change in domain, the model differs from Rumelhart and Norman's typing model in the way in which the gradient of item activations is established. In this case, a static activation gradient is established over word representations via associative connections with a single representation of the sequence to be produced (a 'control element'). Words near the beginning of the sequence have a strong connection to the 'sequence' representation and are thus strongly excited throughout sequence production, while words near the end of the sequence receive only weak excitation and are thus weakly active. The difference in activation from each word to the next is given by a fixed ratio, thus the more active a word, the larger the activation difference between it and its successor. Once a word has been selected for output it is assumed to be in a refractory state for the duration of the sequence and is subsequently excluded from the output competition. Random noise is added to the activation levels of item representations in order to simulate a degree of uncertainty in the competitive output process, and this introduces errors in sequence production.

The model provides a good account for a number of phenomena associated with verbal STM. The errors which occur all involve the movement of items within the sequence, as

the wrong item wins the output competition in a particular position, with exchanges particularly common as in the Rumelhart and Norman's typing model. In fact the modal error type is the exchange of two adjacent items, in good agreement with the empirical data reviewed in Chapter 1. The introduction of random variation in item activation levels limits the span of the model - longer sequences provide more opportunity for errors; additionally, assuming the same high activation level for the first word in a sequence, the longer a sequence the lower the activation levels of items near the end of the sequence, and hence (since the ratio between activations of successive items is fixed) the smaller the difference between each item and its successor. The smaller the activation difference between successive items in relation to the fixed level of random noise, the greater the likelihood of sequencing errors. For the same reason, the probability of error generally increases as sequencing progresses. However, errors are less likely than would be expected in the final one or two positions of the sequence. This is because there are relatively fewer positions into which items near the end of the sequence may move in an error - the approaching end of the sequence effectively limits the possibilities for error (this is similar to the mechanisms for primacy and recency in Johnson's distinctiveness model and Lee and Este's reverberatory model). The resultant serial error curves closely fit those for visually presented sequences in human subjects. Page and Norris (submitted) are also able to account for differences between visual and auditory presentation by making additional assumptions.

The Primacy model provides an important demonstration that a considerable degree of explanatory power may be achieved using a relatively straightforward static activation mechanism. However, Henson (submitted) presents data which suggest that some degree of positional representation is required to provide a full account for verbal STM (this data, showing that protrusion errors follow sequence position relative to start and end, rather than to other sequence items, has already been mentioned as an argument against chaining accounts). The fit which is achieved to serial error curves is impressive. However, in other domains, such as spelling, these curves are more evenly 'U' shaped than is the case in STM, suggesting that the simple short-range end effect responsible for the recency portion of the curve in the Primacy model may not be sufficient to

provide a good model in other modalities. The dynamic activating system models described below provide an alternative approach.

2.7.2 Dynamic models - Low positional resolution

This section discusses a single model - Houghton's CQ model of speech production which uses a dynamic form of activating mechanism but encodes position with respect to a small number of positional reference points. This, together with the use of an activation function of limited positional resolution, places the model more towards the 'ordinal' than the 'positional' end of the position representation continuum.

Houghton's Competitive Queuing model of speech production

Houghton's (1990) model develops the idea of competitive queuing by adding two new features not present in the models discussed above - a dynamic activating mechanism and a learning procedure which enables the model to learn new sequences from exposure. The model is implemented in a connectionist formalism, and has the basic features of CQ systems as outlined above: A set of item representations (in this case the items are phonemes), comprising a set of connectionist nodes, one for each item; an activating mechanism which establishes a (dynamic) activation gradient over the items; and a competitive output system. Houghton defines an explicit mechanism for this competitive component - a 'competitive filter', which consists of a set of nodes connected in one-to-one fashion with the phoneme nodes. The filter nodes are interconnected by mutually inhibitory and self excitatory connections, which establish a winner-take-all competition over them. With the addition of random noise to item representations, the model simulates a number of features of speech production errors, notably the effect of word length, a bowed serial error curve, and the phenomenon of 'co-articulation', with up-coming phonemes primed prior to production.

The dynamically updated activation gradient is the major novelty of Houghton's CQ model compared with its predecessors. As it uses a connectionist approach, the activating mechanism splits naturally into an input signal and an activation function. The input signal comprises two nodes, the 'initiate' or 'I' node and the 'end' or 'E' node, which vary in activity during sequence production. The I node starts with an activation

level of 1.0, and its activation falls exponentially towards 0 as sequence production progresses. The E node takes on a value of 1 minus the I node activation, so its activation is 0 at the start of sequence production and approaches 1.0 asymptotically as the sequence progresses. The activation function is the 'dot-product' rule common in connectionist models - the activating input to a node is the sum of the activations of all nodes feeding into it (the I and E nodes, in this case), weighted by the strengths of the connections to those nodes. While the activation levels of the I and E nodes unambiguously identify the current position in the sequence, the simple dot-product activation rule is able to make only limited use of this positional information. Items near the beginning of the sequence are given strong connections to the I node, and those near the end are strongly connected to the E node. As Houghton (1994a) shows, however, items in the middle of the sequence, which are more weakly associated with both the I and E nodes, cannot be disambiguated by their relative activation levels alone. Thus although this approach gives some of the advantages of a positional representation scheme - in particular, repeated items are possible in a sequence, and the tendency for perseverative errors to occur is reduced - the refractory nature of the item representations (the inhibition of the 'winning' item at each time-step during the sequence) is still crucial for the operation of the model.

2.7.3 Dynamic models - High positional resolution

These models again use a dynamic activating mechanism but are towards the 'positional' end of the continuum, with activating mechanisms which are better able to discriminate successive sequence positions.

Henson's Start-End model of verbal short-term memory

Henson (in press, a.) presents a model of verbal STM, the 'Start-end model' (SEM), which has much in common with Houghton's CQ speech production model, although the implementational details differ and an activation function is used for items which has somewhat better positional discrimination, placing this model in the high positional resolution category.

The SEM is implemented as an abstract mathematical model. (Henson and Burgess, 1997, describe a connectionist implementation of a similar model, using an oscillator-based activation mechanism similar in some respects to that of the OSCAR model, discussed below). Representation of items (words in verbal short-term memory, in this case) take on varying activation levels, and at each time-step the most active is selected for output and then inhibited, as in other CQ models. The major point of difference with other models is in the activating mechanism. Again this falls into two parts, an input signal and an activating function. The input signal at each time-step in the sequence is the distance from the start and from the end of the sequence (by contrast with Houghton, 1990, the end of the sequence is used as a fixed reference point, as well as the start). The activation function has considerably more discriminatory power than the dot-product function, and is able to resolve individual positions in the sequence. The activating mechanism in this model is sufficiently positional that it is able to activate the target item at any point in the sequence to a higher level than its competitors. Strictly speaking the refractory, select-inhibit dynamics of Competitive Queuing are not necessary in this model for the correct generation of the target sequence. Simply selecting the most active item at each stage, without inhibiting it, would be sufficient. However, the model is not totally positional. Items that are close to the target position are also activated by the activating mechanism, although not as strongly as the target item itself. Inhibiting 'used' items, as in other CQ models, thus removes potential competitors and improves the accuracy of the model under noise disruption (in doing this the model takes advantage of the fact that repeats are very much the exception in 'everyday' sequential behaviour). As well as conferring an improvement in accuracy by preventing perseverative errors, the inhibition of past responses is one of the main mechanisms responsible for the model's explanation of the common error types in serial behaviour, as it is in other CQ models.

The model, with the addition of random noise on item representations, provides a good account for many features of verbal STM. The effect of increasing sequence length and the limit on span is similar to other CQ models. The model also gives a good match to empirical data on serial error curves. The increased positional resolution of the model compared with the Primacy model is most critical to its explanation of errors involving

the migration of items from one sequence to a subsequent sequence in the same set of trials (Henson terms these errors 'protrusions'). Henson (submitted) shows that when such errors occur the 'protruded' item tends to end up in the same position in the erroneous sequence, relative to the start and end of the sequence, as it occurred in its original source sequence. The model is able to explain such errors because the positional resolution of the activating mechanism is sufficient to ensure that protrusion of items is likely only when the item's own representation of position matches the current position in the new sequence.

Burgess and Hitch's model of verbal STM

Burgess and Hitch's (1992, 1996) network model of the articulatory loop takes the alternative approach to improving the positional resolution of a CQ model – the standard dot-product connectionist activation function is retained, but the input signal on which it operates is made richer in positional information. Rather than the two dimensions of Houghton's I and E nodes or Henson's start and end positional references, Burgess and Hitch use an input signal with a large number of dimensions – a large set of connectionist nodes, each of which varies in activation independently as sequence generation progresses. By combining the state of each of these nodes the simple dot-product function is able to extract considerably higher resolution information about the current position in the sequence.

The layout of the model is shown in Figure 2.5. A list of words is presented to the model one at a time. As each word arrives, the input phoneme nodes corresponding to its phonemes are activated (simultaneously), and activation feeds forwards from them through a set of pre-learned 'recognition' weights to activate the corresponding word node. A set of temporary weights learns an association between the active word node and the current state of a set of 'context' nodes. The context nodes constitute the high-dimensional input signal discussed above, and the overall pattern of activation over these nodes changes slowly as each word is presented. The same slowly changing sequence of context patterns is repeated during recall of the list, and as each context state appears the temporary weights excite the word node which was associated with the corresponding position in the list at presentation. A layer of nodes with strong

mutually inhibitory and self-excitatory connections form a winner-take-all competitive filter which selects the most active word node. This excites a layer of output phoneme nodes forming the output from the model, and simultaneously inhibits the winning word node.
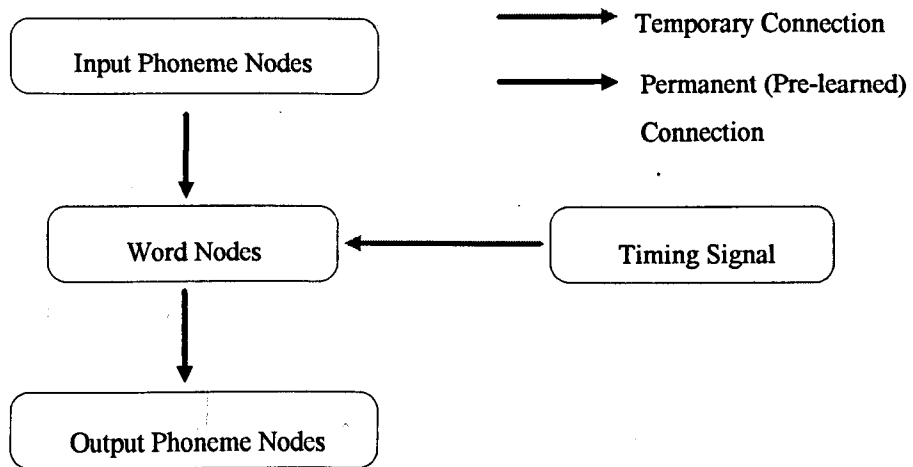
Figure 2.5. *Basic architecture of the Burgess and Hitch model.*

The initial version of the model (Burgess and Hitch, 1992) included the option to use a second set of temporary weights, which learnt an association between the phonemes of one word at the output phoneme nodes and the phonemes of the following word at the input phoneme nodes. These weights form a chaining system which, as a word is recalled, tend to activate its successor. The 'chaining' weights proved to produce types of error incompatible with the human data, and were incapable of producing the type of order error, typified by exchanges, which characterises human performance, while the temporally varying context scheme, by contrast, resulted in performance generally in line with the psychological data. One of the strengths of the CQ approach is that while associative chaining may be accommodated within the general framework, it is not necessary. The poor performance of the Burgess and Hitch model when using chaining reinforces the arguments against chaining approaches elsewhere in this chapter.

Random noise added to word node activation levels, along with temporary weights which are allowed to decay with time, limit the span of the model. The model performs

well in the light of short-term memory data, reproducing the effects of word length and serial position. The primacy portion of the serial error curve is mainly due to the decay of the temporary weights with time, while a short recency effect is produced by a similar mechanism to the Primacy and SEM models - there are fewer possibilities for error in the serial positions close to the start and end of the sequence. (This effect also contributes to the primacy portion of the curve, but is less noticeable due to a ceiling effect). The model also shows the correct predominance of ordering errors, particularly exchanges (when the chaining weights are not used).

The use of a phonemic representation for words at the input and output of the model allows the effect of phonemic similarity to be modelled. Words in the presented list which share phonemes compete more strongly during recall than phonemically distinct words, because words similar to the 'target' word are partially activated during presentation by the phonemic recognition weights and form partial associations with the corresponding temporal context pattern. They are thus partially activated when the same context pattern appears at recall, leading to a greater error rate and hence lower span for lists of phonemically similar words, as for humans. The model's treatment of phonemically similar words will be returned to in Chapter 5, where the operation of serial constraints in CQ models is discussed more generally.

The Burgess and Hitch model has been developed by Glasspool (1995) to treat input list items (both words and nonwords) as a of phonemes, rather than as a parallel set. The extended model proposes that a purely phonological trace is laid down in STM, in addition to a trace containing information about words appearing in the input. The phonological trace alone is used to recall non-words, while it operates in conjunction with the lexical trace to recall word lists. The model can recall sequences of nonwords as well as words, words comprising the same phonemes in different orders, and reproduces the effect of lexical status (the relatively poorer performance on nonwords compared with words). The timing signal for the model has two components, changing at different rates corresponding to the production of phonemes and words. The somewhat similar idea of a composite context signal, incorporating elements entrained to high and low frequency components of the stimulus, is also proposed by Henson and Burgess (1997), Hitch, Burgess, Towse and Culpin (1996) and Brown, Preece and

Hulme (submitted) in connection with grouping effects. Burgess and Hitch (1996) develop the model in a different way, incorporating the ability to deal with nonwords by implementing connection weights with both long and short-term plasticity, and removing the chaining element which disrupts recall in the initial model. This extended model again captures lexicality effects, and the effect of phonemic similarity is more accurately predicted.

OSCAR

OSCAR (OSCillator-based Associative Recall) (Brown, Preece and Hulme, submitted, Vousden, 1996, Brown and Vousden, in press) is a model of sequential memory which has been applied to a number of verbal short- term memory paradigms and to speech production. As with Burgess and Hitch's model, a simple dot-product activation function is used in conjunction with a high-dimensionality input signal, although this signal is of a different type. Items to be learned are associated with consecutive states of this slowly changing temporal 'context' signal, and the same sequence of context states is reinstated during recall, leading to sequential recall of the learned items. Output from the model is by way of a competitive process whereby the most active of a set of potential vocabulary items is selected for output at each step, and is then inhibited to make it unavailable for subsequent recall.

The model is described in terms of vector operations. However, it may easily be interpreted in neural network terms, and the dynamics of its operation are directly comparable to network - based CQ models. The OSCAR model has two major points of novelty in respect to prior CQ models: Firstly, the temporal context which provides the overall timing of the model is based upon a set of oscillators, and secondly the model uses distributed representations for the internal storage of sequence items (the output of the model uses an essentially localist representation; in CQ models this is effectively forced by the fact that select- inhibit output dynamics are used in the output stages).

The temporal context signal, to which the sequencing of items at recall is tied by learned associations, is generated by combining the outputs of a number of sinusoidal oscillators. The oscillators cover a fifteen octave frequency range, and are spaced at

octave intervals - that is, moving through the set of oscillators from fastest to slowest, each operates at twice the frequency of its predecessor. Each of the vectors comprising the context signal thus combines many measures of the current sequence position. However, in contrast to the signal used in Burgess and Hitch's model, the context signal contains elements which repeat at different frequencies. Items to be stored by the model are presented as randomly generated vectors. Each is associated as it arrives with the current state of the temporal context signal. To recall the learned sequence, the context signal is reset to its starting state, and the signal then evolves just as it did during learning, producing a series of context vectors. As each of these is generated it is used to probe the memory trace, resulting in a recalled item vector which is an approximation to the original stored item. As with TODAM, this recovered vector is then 'de-blurred' by selecting the closest matching actual item vector. This winning item is then considered to be inhibited, and is excluded from the competition for subsequent outputs in a process conceptually equivalent to the selection and then inhibition of the most active output item in other CQ models.

Brown, Preece and Hulme demonstrate a number of verbal STM effects, including the separation of item and order memory, the effects of list length and item similarity, judgements of recency, power-law forgetting, serial position effects, and effects of grouping within lists. The use of a temporal context signal with periodic components leads to this latter effect, where items recalled in the wrong group tend to maintain their within-group position. This can happen if a strong periodic component of the context signal becomes entrained to the frequency of grouping in the stimulus list. The context for a particular position in one list then shares features with the context for the same position in another group.

Vousden (1996) applies the OSCAR architecture to speech production, and explains the constraints imposed by syllabic structure on speech errors in terms of the periodic nature of the activating mechanism. Thus for example in exchange errors onset phonemes are more likely to exchange with onset phonemes in other words than with phonemes from other positions within the same word. The periodic self-similarity of the OSCAR input signal offers an explanation for such effects since it is possible to arrange for the input signal to be synchronised with syllable position. Thus the input signal for

an onset phoneme shares features with those of onset phonemes in other syllables. This explanation has similarities with that offered by Hartley and Houghton (1996) in a model of speech production based on a generic, language specific syllable template. The template is cyclic, and generates a periodic template structure to which generation of phonemic output is tied. This model is discussed in more detail in Chapter 5.

2.7.4 Summary

CQ models have been successful in a variety of psychological domains. All three approaches to the representation of sequence position - static, and dynamic with low or high positional resolution - have been employed by modellers with good effect. The static approach allows good modelling of the basic features of verbal STM in the Primacy model, although it does not give a straightforward account for sequences with repeated items and can only produce the short recency region associated with short-term memory rather than the longer 'inverted-U' serial curve of spelling errors, for example. The evidence from 'protrusion' errors indicates a degree of purely positional association, which may lead to problems with modelling more detailed data with static gradient models, even within the verbal STM field. Dynamic position representation with low resolution - Houghton's 'I-E' signal - allows robust sequencing with repeated items and good immunity from perseverative errors and noise (due to the more uniform high activation level of items throughout the sequence), but it requires supervised learning on longer sequences. Higher resolution dynamic approaches generally perform well in their chosen domains and allow robust sequencing with single-shot learning. Both 'select-inhibit' dynamics and positional cueing operate as sequencing mechanisms in these models, which increases their robustness while maintaining the tendency towards ordering and exchange errors. The degree to which each of these mechanisms contributes to the effectiveness of the models as sequence generators is an interesting question which remains open for future work to address empirically. Intuitively, both mechanisms will improve the robustness of the systems by reducing the level of competition from 'incorrect' responses in any serial position.

Another area which promises to be a very interesting one for future work is the degree to which the type of errors to which a CQ system is prone are affected by aspects of the

domain within which it operates. Vousden and Brown (1997) have made an interesting start in this direction with a study of the effect of one parameter in the OSCAR model - the degree of inhibition of items following their production - which seems to be the key to some of the differences between the types of error common in speech as opposed to short-term memory. Increasing the level of inhibition causes the pattern of errors produced by the model to change from one similar to that seen in speech errors to one more typical of short-term memory experiments. Vousden and Brown relate this parameter difference to different task demands in the two areas. Another factor which may be important is the nature and degree of serial 'category constraints' in different domains - for example the constraint on which category of phoneme may occur in particular syllabic positions. The incorporation of this type of constraint into CQ systems is a particular focus this thesis and Chapters 5, 6 and 7 will explore the matter in some detail.

Finally it is interesting to note that a model which takes a similar approach to the generation of sequential behaviour - the production of a dynamic gradient of activations and the selection and inhibition of the most active response - has been successful in a rather different area of psychology. The Norman and Shallice (1980, 1982) model of willed and automatic control of behaviour posits a two-component view of the control of action. Over-learned or habitual action is held to be controlled by a set of schemas competing within a contention scheduling (CS) system for control of the motor system, while willed or attentional control of action is achieved by a supervisory attentional system which influences the CS system. Cooper and Shallice (in submission, 1997), model the contention scheduling element, which comprises a hierarchy of schemas terminating in a set of low-level actions which are carried out directly by motor systems. Schemas are activated from the top down by their parent schemas or willed control, and from the bottom up by input from the environment. They compete for execution on the basis of their activation level. A schema is triggered when its activation level is higher than any other schema and higher than a trigger threshold. A triggered schema feeds activation forward to its child schemas, and is inhibited after its goal has been achieved. Although this model operates at a higher notional level than the other models discussed here, the means by which serial ordering of actions is achieved

is basically that of CQ. The approach is particularly interesting because it is motivated by ideas concerning the evolution of automatic and of willed behaviour systems, and may offer an independent rationale for the processes underlying CQ.

The models discussed above show that the CQ approach has been successful in modelling empirical data from a wide range of serial domains. Importantly, CQ models are able to explain the constellation of features of serial behaviour highlighted in Chapter 1. This sheds new light on the question raised at the end of that chapter - does the occurrence of this set of overlapping features of serial behaviour imply shared processing subsystems or separate subsystems which all operate in a similar manner? If the features represented the chance co-occurrence of a number of arbitrary mechanisms, they would be unlikely to occur in more than one processing system, so the shared system hypothesis would be greatly favoured. However, any mechanism which uses the CQ approach is likely to exhibit several, if not all, of these common features. The question remains an empirical one, of course, but the existence of a computational approach to sequence generation which naturally exhibits this common set of error features when it breaks down adds significant weight to the argument for the separate subsystems hypothesis.

3

The spelling system and graphemic Buffer Disorder

The theoretical points to be discussed in the remainder of this thesis will be illustrated through a number of computational models in a single domain, that of spelling. Two features of spelling make it an interesting area for the development of theoretical approaches to serial behaviour. Firstly, spelling is a serial behaviour which is easy to study, and much experimental data is thus available to motivate theoretical accounts. Secondly, a particular acquired deficit to spelling exists - so-called 'Graphemic Buffer Disorder' (GBD) - which appears to mainly concern the task of generating the correct sequence of letters, independent of more general aspects of linguistic processing. The study of this disorder is thus particularly interesting from the point of view of understanding serial processes in general. A major motivation for the computational models presented in later chapters is to gain an understanding of the deficit in GBD patients. Some features of the syndrome have already been described in Chapter 1, but in order to properly situate the modelling work it will be useful to give an outline of the spelling system in general and to describe GBD in more detail.

3.1 A theoretical framework for the spelling system

Spelling is an acquired skill of fairly recent origin in evolutionary terms, and it does not therefore seem likely that it is subserved by any special purpose brain mechanisms. More probably it is parasitic on pre-existing linguistic and sequential control systems. It is of considerable interest, therefore, that certain neurological syndromes (acquired agraphias) exist in which aspects of spelling may be selectively disrupted. Two types of deficit in particular have been used as evidence for the gross structure of the spelling system:

1. Certain patients ("phonological" agraphics) show impaired spelling for novel dictated nonwords, while their ability to spell words already known to them is preserved. For example, patient PR of Shallice (1981) achieved 91% correct spelling

for known words, but only 18% correct spelling for nonwords. The preserved spelling of real words extends even to irregular words like 'yacht'.

2. Conversely, "lexical" agraphics have impaired spelling for words together with preserved spelling for nonwords. Such patients produce regularisation errors on words (e.g. *yacht* → 'yot') and their spelling for known words is worse the more irregular the word is, suggesting that they are spelling only by exploiting sound-to-spelling regularities. For example, patient RG of Beauvois and Derouesne (1981) produced perfectly reasonable spellings for nonwords, was 93% correct spelling regular words, but only 36% correct when spelling exception words.

The two deficits form, in Shallice's (1988) terms, a 'strong' double dissociation, implying that the two tasks involved - spelling known words on one hand and spelling nonwords and regular words on the other - rely on different subsystems. This evidence, along with common-sense reasoning about the types of process which might be involved in spelling a word, has been the major motivation behind the now standard view of the spelling system as comprising dual routes. Figure 3.1 shows the typical form of such functional models, and is similar to several 'dual route' formulations in the literature (for example Morton, 1980, Seymour and Porpodas, 1980, Shallice, 1988, Link and Caramazza, 1994). Here, route I is a lexically based system, capable of spelling known words using learned spellings from a graphemic lexicon, and route II is an assembled route, using procedures to generate spellings from the phonological form of the target word. The lexical route is able to retrieve the spellings for both regular and irregular words once they have been learned, but it is not capable of spelling a novel word on first exposure, when no lexical entry exists for that word. The assembled route is able to construct a spelling for any word online using a set of procedures and rules which translate from the phonological form of the word to a spelling. The success of this route (i.e. how closely the assembled spelling resembles the 'correct' received spelling for the target word) depends on how successfully the mapping from phonology to orthography for that particular word is captured by the procedures available. A range of different types of procedures could be employed by this route (see Glasspool, Houghton and Shallice, 1994, for a discussion), but in general the mapping will be more successful for regular words (which by definition have a phonology-to-orthography

mapping which conforms to the norms of the language) than for irregular words. However, this route is capable of supplying a candidate spelling for a novel word. On this framework, phonological agraphia would correspond to a deficit on route II, and lexical agraphia to a deficit on route I.
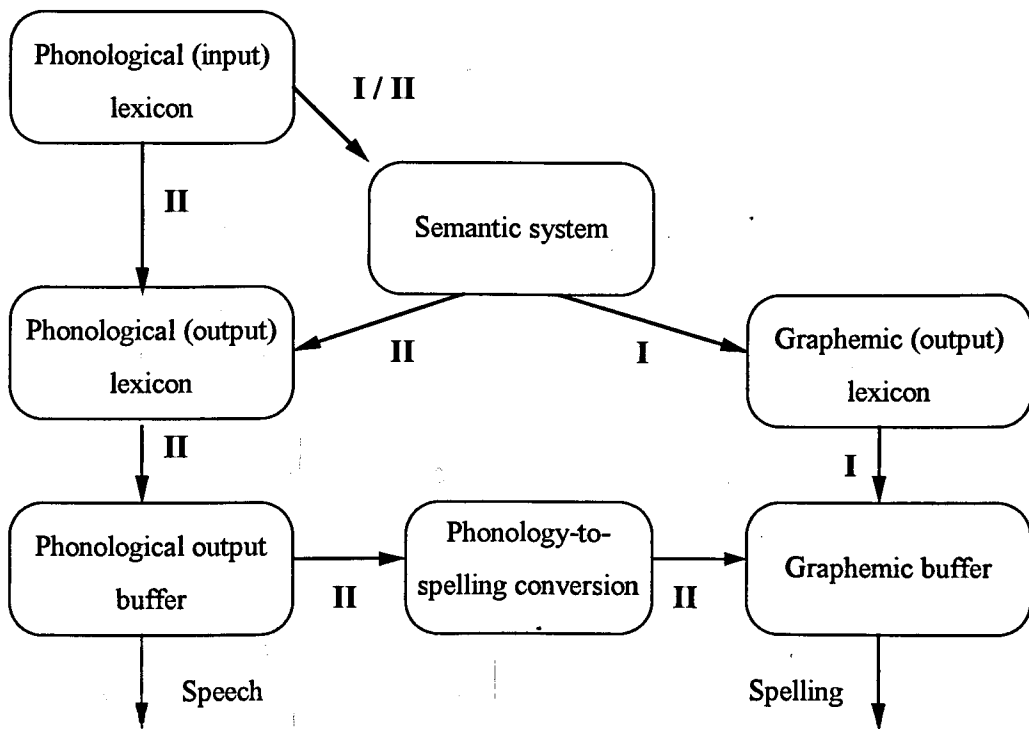
Figure 3.1. *A functional diagram of the human spelling system.*

3.2 Graphemic Buffer Disorder

The fractionation of the spelling process into two routes implies that they come together again at some point in a final common path. The point of confluence is generally held to be a buffer - the "Graphemic Buffer" of Figure 3.1 - which can hold the final spelling while subsequent processes specific to the modality required for output (e.g. written or spoken spelling) operate on it. (Ellis, 1979; 1982, has developed a comprehensive model of the processes involved in the realisation of spellings from the modality-neutral representation assumed to exist at the level of the graphemic buffer).

What type of deficit would result from damage at this point in the spelling system, to the graphemic buffer itself? Assuming that the damage was sufficiently mild that

spelling were still possible at all, one would expect that the lexical status of the stimulus (word or nonword) and, if a word, factors which might affect lexical access (such as frequency or grammatical class) would not have a qualitative effect on performance, and that errors due to the operation of the inappropriate spelling route (such as regularisations) would not occur. However, the nature of the difficulty beyond these generalisations cannot be deduced from the functional model. It will depend critically on the mode of operation of the 'buffer' component itself, and possibly on the type of damage sustained.

At least nine agraphic patients have been described who fit this general pattern: FV (Miceli, Silveri and Caramazza, 1985), LB (Caramazza, Miceli, Villa and Romani., 1987, investigated in more detail by Caramazza and Miceli, 1990), SE (Posteraro, Zinelli and Mazzucchi, 1988), DH, ML (Hillis and Caramazza, 1989), HR (Katz, 1991), HE (McCloskey, Badecker, Goodman-Shulman and Aliminosa, 1994), JH (Kay and Hanley, 1994), CW (Cubelli, 1991) and AS (Jonsdottir, Shallice and Wise, 1996). The deficit in these patients appears limited to a difficulty with the sequential selection of letters, higher and lower levels of the spelling processes being preserved. Thus the errors affect all relevant modalities of output, in particular writing and spelling aloud, and affect word and non-word spelling similarly. Furthermore the errors of these patients do not show the patterns typical of either phonological or lexical agraphics - in particular, neither phonological nor semantic paraphasias are common (in fact, errors often result in letter strings which are unpronounceable), and in most cases errors show little effect of word frequency or class. This suggests that the deficit is not a result of simultaneous damage to both routes (either of which would be expected to show these characteristics when damaged) but is located instead within the shared stages of the spelling process.

If the locus of damage in GBD patients is indeed the graphemic buffer, then the pattern of errors made by these patients becomes very interesting from the point of view of understanding serial behaviour. The buffer represents the point at which a representation of the spelling is assembled for serial read-out, and the disorder may thus shed light on the process of serial output itself. As was briefly discussed in Chapter 1, the spelling errors made by GBD patients show a distinctive pattern. While the studies

of different patients vary greatly in the level of detail available, the consistent features of the disorder are as follows:

Effect of word length

There is a strong effect of word length on the patients' ability to write a word correctly, with all patients showing worse performance on longer words. The magnitude of this effect varies, however. Patient JH, for example, with a particularly mild impairment, shows a reduction in performance from 90% correct on three-letter words to 70% on seven-letter words, whereas patient AS shows a performance reduction from 96% to 44% over the same range.

Error types

The majority of the errors can be explained in terms of a single substitution, deletion, transposition or insertion of letters. Different patients vary somewhat in the relative proportions of these different error types, however. ML, HR, DH and SE show a predominance of deletions (in fact almost all of HR's errors are deletions), while FV, for example, produces relatively few deletions (10%)[1]. For LB, FV, CW and JH substitutions are the most common error (64% of FV's errors are substitutions, for example). HE and AS produce similar numbers of substitutions and deletions. Most patients also produce a fair number of insertion and transposition errors. For example, LB's errors include 6% insertions and 17% transpositions, while for AS the proportions are 22% and 14% respectively. For those patients where the distinction has been made (LB, JH and AS), transposition errors are predominantly exchanges rather than shifts (shift errors were very infrequent for these patients). Mixed errors, containing more than one of the error types listed above, also occur, with different incidences in different patients although such errors are generally at least as frequent as any of the individual error types.

---

[1] Unless otherwise stated, all percentages are of responses showing only a single type of error.

The effect of word length on the relative proportions of different error types is reported for AS, LB, JH, DH, and ML. Comparing the influence of word length across these patients a pattern emerges, with AS, LB, JH and DH all showing an increase in the relative proportion of deletion errors and a decrease in the proportion of substitutions with increasing word length. ML however does not appear to show any trend in these error types, and the patients are variable with respect to the effect of word length on other error types.

Effect of consonant/vowel status

In those patients where the matter has been reported, there is a strong tendency for exchange and transposition errors to preserve the consonant or vowel (CV) status of letters. Thus errors on consonants tend to be consonants, and vowels tend to be replaced by vowels. The effect is noted for patients LB, AS, JH, HE, FV and SE. The effect is not absolute; thus for AS, for example, 82% of substitution and 62% of transposition errors preserved CV status. LB's errors show a much higher rate of CV preservation - 99% of substitutions and 91% of exchanges preserve CV status. Jonsdottir, Shallice and Wise (1996) interpret this difference as reflecting the fact that AS is English whereas LB is Italian. Italian has a considerably more regular phonology-to-spelling mapping than English, and this could affect both the mechanisms and internal representations adopted by the spelling system. An Italian GBD patient may be able to use phonological information more effectively than an English one to constrain possible errors, or CV status may be a more salient feature of letter identities in Italian than in English.

Double letters

Patients show an interesting range of errors on double (or geminate) letters. In errors the property of gemination (i.e. doubling of a letter) appears to separate from the identity of the letter being doubled. This feature has been noted in the errors of patients LB, AS and HE, and is apparent in several types of error. For example, LB produces errors where the doubling occurs in the wrong position (e.g. sorella → sorrela), the doubling does not occur (sorella → sorela), the doubling occurs in the correct position

but with the wrong letter doubled (sorella → soretta), or where an extra doubling occurs (sorella → sorrella). LB virtually never introduces a double letter in a word which does not already include a doubling. Patient AS makes similar types of errors, and again the introduction of a doubling into a word without one is very rare.

Serial position effects

For most patients errors are more common in medial letter positions than at the start or end of a word. The exceptions are FV, who shows no effect of serial position, and HR, who shows a monotonic increase in error probability from start to end of word. Among the majority of patients exhibiting bowed serial error incidence curves the mode of the distribution varies. Thus LB and ML show a peak in errors nearer the start of words, whereas AS and DH show a peak towards the end. Other patients show more symmetrical serial error curves.

Effect of lexicality

It is part of the definition of GBD that errors on words and nonwords should be qualitatively similar, and this is the case for all patients who have been tested in sufficient detail (LB, FV, AS, JH, SE). However, only SE shows quantitatively similar performance irrespective of lexical status. All other patients show a quantitatively poorer level of performance on nonwords compared with words. While a difference in performance between nonwords and words is not directly predicted by the functional framework of Figure 3.1, it is compatible with it and could correspond for example to a lower level of excitation to the graphemic buffer for nonword compared with word input.

Effects of syllabic structure

An effect of syllable complexity has been described by Caramazza and Miceli (1990) in the Italian patient LB; thus simple structures, ones composed entirely of CV pairs (e.g. Milano), were better spelled than complex ones (e.g. Stresa) and showed a somewhat different pattern of errors. However the effect does not occur in the English patients JH and AS. Jonsdottir, Shallice and Wise (1996) have suggested that the more transparent

phonological to orthographical mapping of Italian may enable high-level phonological information such as syllable structure to be used more effectively in spelling.

3.3 Comparison with normal spelling

It is interesting to note that several features of GBD are also present in the 'slips of the pen' of normal spellers (Wing and Baddeley, 1980, Hotoph, 1980). The same types of error occur, albeit at much lower incidences, and similar effects of word length and serial position are found. This supports the position that the errors of GBD patients are due to the degraded operation of the normal spelling system, rather than, for instance, interference from other processes. There are however some differences. GBD patients tend to make more exchange errors and fewer deletions, relatively, than normal spellers, and the preservation of CV status seen in some GBD patients has not yet been observed in slips of the pen. This thesis will concentrate on modelling GBD errors rather than slips of the pen, and will treat GBD as a unitary disorder of sequencing in spelling. The issue of other processes in spelling will be returned to in Chapter 8.

3.4 Modelling GBD

The ^{SEVEN} error features listed above are not explained by the localisation of these patients' deficit to the Graphemic Buffer, but presumably reflect the internal operation of that component. How are these more detailed characteristics of the disorder to be explained?

Caramazza and Miceli (1990) have proposed a model - the multiple object spelling ("MOS") model (see also Link and Caramazza, 1994; Tainturier and Caramazza, 1996) - concerning the nature of the representations in the graphemic buffer. They take the view that patterns in GBD errors directly reflect points of vulnerability in the internal representations employed in the 'buffer', and propose an abstract structure for these representations. On the model, multi-dimensional graphemic representations exist in the graphemic buffer, with separate 'tiers' corresponding to different classes of structural information (Figure 3.2). A set of principles govern the way the data structure is used. The orthographic representation of a word is held to comprise a set of (ordered) letter identity tokens, each bound with a C or V token corresponding to the letter's

consonant/vowel status, and optionally bound with a geminate token indicating that the associated letter should be doubled. The letter tokens are organised by bindings to higher level elements identified as 'graphosyllables' - hypothetical orthographic equivalents to phonological syllables, differing in some cases in the location of their boundaries.

Orthographic Syllable Tier

Consonant/Vowel Status Tier

Letter Identity Tier

Geminate Status Tier
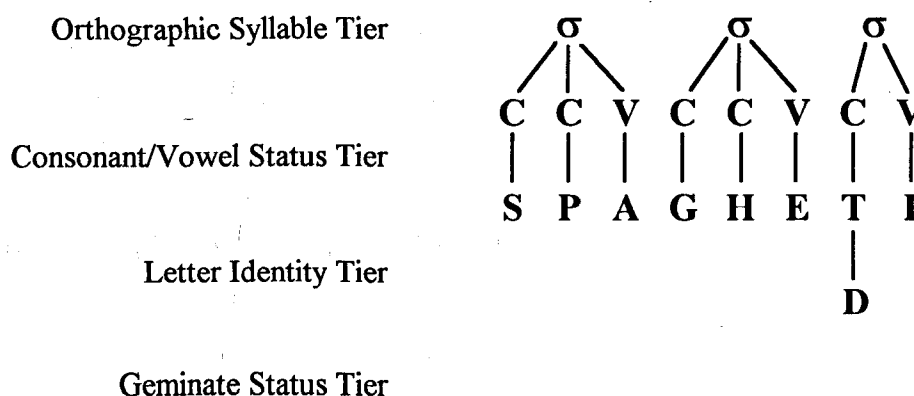
Figure 3.2. *Representation of the orthographic structure of the word spaghetti according to the MOS model of Caramazza and Miceli (1990). Words are divided into orthographic 'syllables', and tokens representing letter identity, consonant or vowel status and the doubling of a letter are bound into the representational structure at different 'tiers'.*

Four principles, either explicitly stated or implicit in Caramazza and Miceli's discussion of the theory, govern the production of erroneous spellings by a GBD patient:

1. Damage to the spelling system is assumed to manifest itself as the (random) loss of certain information from the 'tiered' orthographic representations - either letter identities or bindings.

2. Other than at points where the representation has been damaged, the binding between the tokens on each of the four dimensions is assumed to hold, so that, for example, a letter bound to a geminate token may "break off" from their associated CV structure and graphosyllabic tokens, and move as a pair to re-attach at some other point on the structure.

3. It is assumed that the CV tokens constrain possible bindings. Thus a consonant letter token will bind preferentially to a consonant CV structure token, and a vowel letter to a vowel CV token.

4. Finally, in order to explain the blocking of certain errors, a "Minimum Complexity Principle" is assumed. This is held to constrain the response produced when information is no longer present in the buffer to "the least complex graphosyllabic sequence consistent with the available information". An ordering scheme is assumed, by analogy with phonological theory, by which the simplest possible structure is the syllable composed of a consonant as onset and vowel as rime (CV).

The model assumes that the structural links between information have different strengths, and thus degrade more or less readily when subject to damage. The structure of the representation, and the relative strengths of the links, are based on the errors characteristic of GBD. The theory thus stated is claimed to account for the major features of the data: The dissociation of orthographic from phonemic effects, the preferential substitution of vowel for vowel and consonant for consonant, and the behaviour of geminate clusters. The "Minimum Complexity Principle" also provides an account for the blocking of such errors as the deletion of a singleton vowel or consonant (V in CVC or C in VCV). These errors result in representations which are more 'complex' than the original stimulus, and are thus blocked by the Principle.

The MOS approach does not aim to comprehensively model the serial processes of spelling. No algorithm is specified to operate on the structured representation, without which quantitative predictions cannot be made, and even qualitative results rely on intuition about the way in which representations might be likely to break down. Partly as a result of this the approach is open to criticism on the grounds that it simply describes the data rather than explaining it. The various features of the proposed representation are included purely to account for the data and have no independent, principled motivation. Additionally, the model is not comprehensive enough to account for a number of salient patterns in the data - the effects of serial position and word length and the relative incidence of different types of error, effects at least as strong as those which the model does address.

Some aspects of the GBD error pattern, in particular the effects of serial position and word length, are strongly reminiscent of the breakdown in sequencing in CQ models. This suggests that the CQ approach might be a fruitful one for a concrete performance model of the graphemic buffer. Taking this observation as a starting point the following chapters set out to develop just such a model. The serial position and word length effects and the nature of individual errors are the features most obviously amenable to modelling using CQ since similar effects have been demonstrated in previous models. Taking an incremental approach the initial model will address these general features, with subsequent modelling building further spelling-specific features onto the resulting substrate.

3.5 Modelling phonological spelling

A number of computational models have been proposed for spelling which include both the phonological and lexical routes. Brown and Loosemore (1994) describe a connectionist model which uses a backpropagation procedure to learn a mapping between a representation of the phonology of a word and a representation of its spelling. The model is able to generalise its knowledge of sound-to-spelling correspondences to produce reasonable spellings for non-words, and is able to model aspects of developmental dyslexia when its computational resources are reduced (by reducing the number of hidden units). The model does not aim to account for slips of the pen or for serial order errors in spelling, and uses a static rather than dynamic representation for phonology and orthography. Olson and Caramazza (1994) again model the mapping from phonological input to orthographic output in a connectionist network, but use a serial representation for input and output. The input is a phonemic representation of the target word which is shifted one position at a time across a field of phoneme positions at the input to the network. The output is a sequence of letter representations. The model does not generate serial order endogenously but performs a running translation from serial input to serial output. As with the Brown and Loosemore model the network learns to map from sound to spelling. However, serial order errors typical of slips of the pen or GBD are not reported for the model. Glasspool, Houghton and Shallice (1995) describe a somewhat similar model which uses a less powerful learning procedure and captures only the sound-to-spelling

correspondence aspect without supporting lexical spelling. Since the problem of serial order generation, rather than the spelling process in its entirety, is the focus for this thesis, the phonological-to-orthographic mapping will not be pursued here and the modelling work will be limited to the serial output processes in spelling.

4

CQS: An initial CQ model of GBD[2]

GBD is a useful target disorder for computational modelling from the point of view of elaborating a general theory of sequential behaviour, essentially for two reasons:

1. The disorder appears to be limited to the final stages of the spelling process involving the generation of a serially ordered output. As such it promises to speak directly to the problem of serial order in behaviour.

2. Those gross features of the disorder which are not directly accounted for by the localisation of the deficit in the "graphemic buffer" are strongly suggestive of the forms of error typical of Competitive Queuing models. A model taking the CQ approach thus offers the promise of a principled explanation for these features of GBD, and if successful the 'finer' level aspects of the error pattern should provide a good test of the predictive power of the approach.

As a first step towards this goal this chapter presents an initial Competitive Queuing model of the serial output stage of the spelling process.

4.1 Scope

The strategy adopted here towards modelling the spelling process will be to start by aiming to account for those features of normal spelling errors and GBD which are part of the established performance of CQ models - from the point of view of CQ modelling

---

[2] The model and some of the simulations in this chapter have been reported in three published papers (Houghton, Glasspool & Shallice, 1994, Glasspool, Houghton and Shallice, 1994, Shallice, Glasspool and Houghton, 1995). This chapter discusses the model in greater depth than these papers, and includes a number of new and previously unreported simulations.

these might be termed the 'core' features of GBD. This model is to provide a base upon which to build more detailed models, so accurate quantitative modelling and detailed aspects of the data will be considered less important than overall qualitative performance. Specifically, for this initial model the aim is to provide a qualitative account for:

1. Word length effects.

2. Serial error curves.

3. The occurrence of insertion, deletion, shift, exchange and substitution errors.

The CQ approach demands some form of special treatment for doublings, so in order to allow the model to handle words containing double letters some additional special-purpose mechanism will be required. It is hoped that the typical error pattern on geminate letters might be modelled by disrupting a straightforward geminate mechanism. A subsidiary aim is thus to:

4. Account for the types of error made by GBD patients on double (geminate) letters.

These four effects do not entail the differential treatment of different classes of letters. The effects of consonant/vowel status of letters will require differentiation at least between consonant and vowel letters and this will be addressed in later chapters.

4.2 Approach

Since the 'core' features of GBD are so suggestive of the typical error behaviour of a generic CQ model, the approach taken in this model is to use as basic a CQ mechanism as possible, the idea being to see how far it is possible to get in a simulation of GBD spelling using a generic sequential behaviour system with few if any spelling-specific modifications. The instantiation of a theoretical system as a concrete computational model always requires a number of design decisions to be made, which fall into two types: Concrete implementations must be chosen for theoretical elements which may previously have been specified in a more abstract way, and precise numerical values must be chosen for parameters which scale the interactions between these elements.

Free parameters can be tackled in a number of ways, which will be explored later in this and succeeding chapters. The 'higher level' type of design decisions must be made before modelling commences, however. The guiding principle for this model is the desire to keep the system both as simple and as general (i.e. non spelling-specific) as possible. With this in mind, the following points detail and justify the decisions made concerning the overall architecture, and thus lay out the bare bones of the model which will be fleshed out in the next section. Since the modelling decisions made here also form the point of departure for the design of the models of later chapters, they will be · justified in some detail.

4.2.1 Formalism

The 'generic' formulation of the CQ approach presented in Chapter 2 (sections 2.5 and 2.7) does not by its nature imply any particular type of implementation. However, since it centrally involves associations between items with variable activations a connectionist formalism is convenient for its implementation (most implemented CQ models are connectionist in nature). The basic CQ model of Houghton (1990), which employs a "localist" connectionist formalism and is straightforward in its architecture, is thus taken as a starting point. The simplest way to adapt this model, which was targeted at the domain of speech production, to operate in the domain of spelling is to interpret the 'item' nodes as letters rather than the phonemes of the original.

4.2.2 Timing signal

Since repeated letters are common in English spelling, the model will require a dynamic rather than static timing signal. The simplest such signal which has been used previously is Houghton's (1990) 'I-E', or "start-end' timing signal, in combination with a simple 'dot-product' activation rule for item nodes. This approach provides more information about position in sequence than does a static timing signal (e.g. Rumelhart and Norman, 1982, or Page and Norris, submitted), but less than context signals with greater numbers of elements (Burgess and Hitch, 1992, 1996, Glasspool, 1995) or with more powerful activation functions (Houghton, 1994a).

4.2.3 Training

Having chosen the 'I-E' type of activating signal a number of consequences follow. Importantly, as a result of the limited power of the signal longer sequences will require a period of supervised training for correct performance. However, the necessary training procedure is already established (Houghton, 1990), and since spelling is an acquired skill which requires an extended period of learning the inclusion of a training phase in the model does not seem an unreasonable limitation.

4.2.4 Geminate (double) letters

As discussed in Chapter 2, the basic CQ architecture is unable to support immediate repeats. Double letters are common in English spelling, and the model will thus require additional mechanisms to handle them. The gross pattern of errors on geminates in GBD patients is similar to that seen in typing errors, and Rumelhart and Norman's (1982) typing model successfully modelled such errors in that domain using a doubling 'schema' which triggered at the appropriate point in word production and served to double whatever letter was being produced at that moment. It would be difficult to conceive a simpler mechanism to address the problem of double letters, so the same general approach is adopted in this model. A mechanism to actually insert the doubled letter into the output stream is also required, but this does not appear to be part of the system directly involved in GBD (in patient LB, performance on six-letter words containing one doubled letter is very similar to performance on five letter words with no doubles. This suggests that words containing double letters are internally represented as a string of letters without the doubling, plus a separate representation of the geminate feature. If the doubled letter is explicitly introduced into the sequencing system responsible for the GBD syndrome one would not expect this performance difference to be possible - six letter geminate and non-geminate words would both contain the same number of letters as far as the damaged system was concerned and would thus both be equally susceptible to errors. This suggests that in the system we are concerned with geminate letters are treated as a single letter, but one flagged as a geminate.) The actual doubling of the letter can thus be assumed to happen external to the system being modelled here. The training scheme of Houghton (1990) must be adapted to allow the

geminate production system to learn when and where to trigger the geminate mechanism.

4.2.5 Stopping strategies

When noise is present in the output stages of a CQ model it can go on producing strings of output items ad infinitum, as there will always be a winning item activation however low it might be. One must therefore consider strategies for halting sequence production at the appropriate point. There are three strategies which lend themselves to implementation in a CQ system:

a) Stop when the correct number of items has been produced. This would suggest the use of some kind of efferent copy, or template, for the sequence specifying the required number of items but not their identity. However, many errors produced by GBD patients are insertions or deletions, where the number of letters in the response differs from that in the target word. Stopping after the correct number of letters has been produced would block this type of error. Moreover, the CQ approach provides a very natural account for insertion and deletion errors whereby items 'cascade' following the error into their correct *relative* positions, although these may not be correct in absolute terms. An approach that, for example, allowed the efferent copy to be noisy so that words occasionally were produced of the wrong length would miss out on this natural explanation for insertion and deletion errors.

b) Stop when no output item is strongly active. This is perhaps the most natural way to stop a CQ sequence. Failure of any item to reach a threshold would imply that all items which should be in the sequence (or at least all those close in the sequence to the current position) have been produced and inhibited, and those that remain are receiving activation from random noise alone, not from the timing signal. Unfortunately, this strategy becomes difficult to pursue when the system must produce item errors (that is, erroneous items which do not appear elsewhere in the target sequence). Such errors are fairly common in GBD spelling, and the obvious way that they would occur in a CQ model would be for the noise level to be set high enough that occasionally noise alone is sufficient to activate letters from outside the sequence, with no excitation from the control signal, to a level where they can

compete with items within the sequence. However, with this level of noise it is not possible for a simple threshold to distinguish between sequence and non-sequence items. If the noise is set high enough to allow item errors a large proportion of non-sequence items will exceed the threshold due to noise alone. With noisy output activations a stop threshold is thus not a practical proposition.

c) Stop when a specific item (a stop 'symbol') is generated. This strategy would allow insertions and deletions to occur in the natural way for a CQ framework, and would not be susceptible to the same disruption under noise as the threshold option. The only danger would be the insertion or deletion of the stop symbol itself, assuming that this is treated simply as another sequence item. However, by definition this item would only appear at the end of the sequence, which should be relatively immune from error on the CQ approach.

On balance, then, the 'stop symbol' approach seems the best option for this initial spelling model.

4.2.6 Differentiation of output items

As discussed above, in order to keep this initial model simple no distinction will be made between consonant and vowel letters. The strategy is to produce a model of the 'core' disorder to which CV distinction may be added at a later stage in order to extend the account to those features of GBD spelling which appear to depend on this distinction.

4.2.7 Integration into the overall spelling model

Chapter 3 set out an overall structural model of the spelling system. How would the current model fit into that scheme? The CQ approach treats the notion of a 'buffer' as an abstraction - there is nothing in a CQ model which corresponds physically to a buffer temporarily holding a set of responses. The sequence of responses made by the system arises from the interaction of individual elements (letters in this case) which are well learned and presumed to be permanently available. This being the case, the model outlined here would best fit into the scheme of Chapter 3 as an implementation of the

lexical route - the system which generates spellings of known words from long-term memory. The 'graphemic buffer' itself would disappear, being on this account an abstraction of the process of sequence generation in the lexical route rather than a physical entity.

The identification of the current model with the lexical route leaves a problem however - how is the output of the assembled route to be integrated into the final output of the system? One solution is to assume that the current model, as well as having the ability to form long-term representations for familiar words, can also rapidly learn temporary representations for the spellings of novel words. Assuming this structure, on the general spelling framework introduced in Chapter 3 the assembled route may be seen as an alternative means of supplying spellings to be learned by this system, and the process of spelling a nonword would involve the generation of a spelling by the assembled route, the rapid generation of a temporary representation for that spelling by the CQ system, followed by the output of that spelling in exactly the same manner as for a known word (this approach is taken by Glasspool, Houghton and Shallice, 1994). This allows a candidate explanation for the difference in performance, but not qualitative pattern, of GBD patients on words and nonwords: Words may be over-learned through an extended training period, whereas on nonwords the model may be trained for the minimum period necessary for correct reproduction. Words may thus acquire a more robust representation than nonwords in the face of a similar level of disruption.

4.2.8 Modelling damage to the spelling system

It has been argued elsewhere (e.g. Houghton, Glasspool and Shallice, 1994) that one obvious manipulation exists whereby non-specific damage to the operation of a CQ system may be simulated - the addition of random noise to the activation levels of competing items. This has the effect of rendering the competitive process non-deterministic, and corresponds to a loss of positional definition in the sequencing process. Real neural systems are subject to continual random noise, due both to the intrinsically noisy characteristics of neurons themselves and to the essentially unpredictable effects of crosstalk from processing in unrelated systems. A decrease in signal to noise ratio in the competitive part of any real CQ-like system is a likely

consequence of any damage which disrupts the flow of activation between elements or the sensitivity of elements, wherever it may occur within the system. Such a manipulation makes the minimum of assumptions about which part of the system is damaged and in what way. Moreover, in a localist model where direct 'lesioning' of nodes or connections is likely to have catastrophic and very specific effects the use of noise is a straightforward way to achieve a non-specific degradation in performance. Accordingly, the addition of a random element to the activation levels of competing items is the approach taken to simulating damage in the present model.

4.3 Architecture and operation

4.3.1 Overview

This section will describe in detail the structure and operation of the model, which will be referred to as the CQS model[3]. The model consists centrally of three layers of connectionist nodes (Figure 4.1). Each node may take on activation levels between -1 and 1, with a resting activation of 0. Only positive activations propagate along the weighted connections between nodes. The first layer consists of pairs of nodes, each comprising an I ('initiate') and E ('end') node. These generate the dynamic timing signal for the spelling process, and each word is represented at this layer by one I-E pair. The nodes operate in the general manner outlined in Chapter 2, section 2.7.2 - at the start of the word, the I-node has a high activation, and this falls as spelling of the word progresses. The E-node has qualitatively the complimentary behaviour, starting with a low activation and finishing up highly active at the end of the word. While the model presumes that each word in its vocabulary is represented by a separate I-E pair, along with the associated layer 1 to layer 2 weighted connections, in the following discussion it will often be convenient to assume that layer 1 contains only a single I-E pair, representing a single word. An assumption of the model is that all connections are unidirectional, so there can be no interaction between I-E pairs at the layer 1 level. This simplification does not therefore affect the predictions of the model.

---

[3] For 'CQ Spelling'.
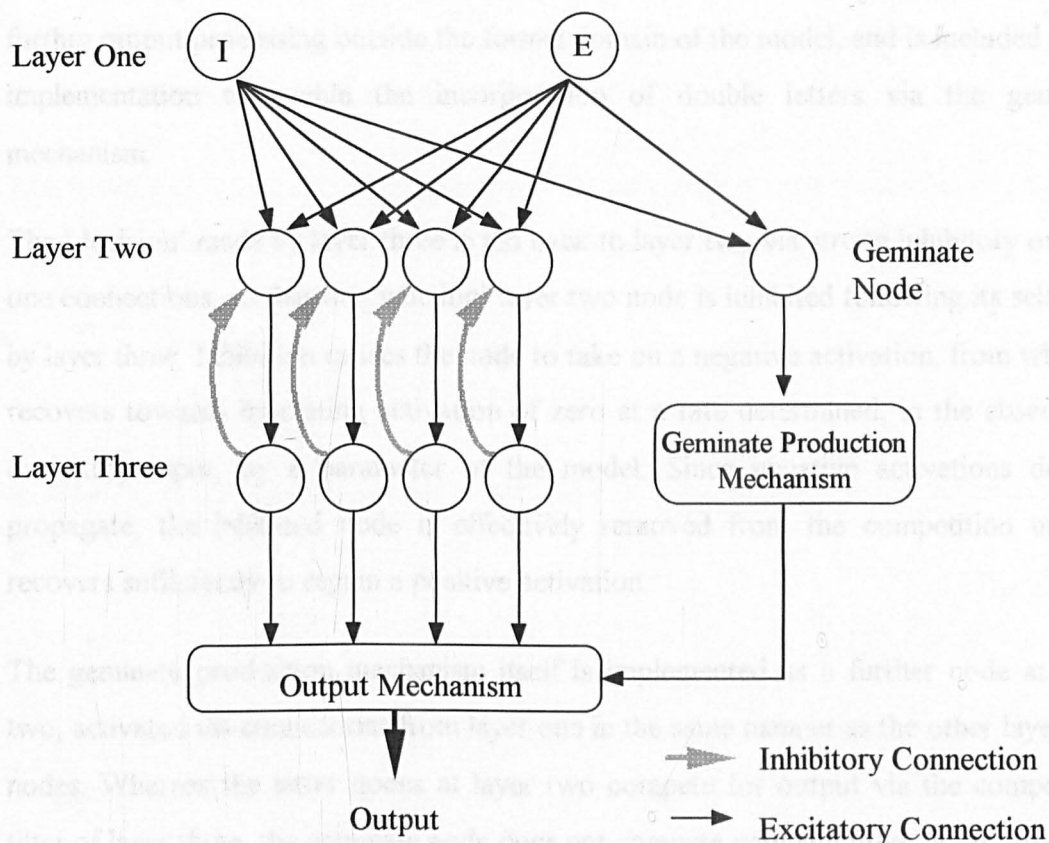
Figure 4.1. *Architecture of the CQS model. See text for details.*

The second layer comprises a set of 26 letter nodes. These are activated via weighted connections from the I and E nodes of the first layer in such a way that a gradient of activations is set up over the letters of the target word, in the usual CQ manner. The nodes are interconnected with mutually inhibitory and self-excitatory connections to form an 'on-centre, off-surround' contrast enhancement arrangement. The weights on these connections are relatively small, however.

The third layer again consists of 26 nodes, each of which is activated via a one-to-one connection from the corresponding letter node at layer two. The layer three nodes are interconnected with mutually inhibitory and self-excitatory connections just as the layer two nodes are. However, in this case these within-layer connections are strong, and cause the layer to operate as a winner-take-all 'competitive filter' (Houghton, 1990). The node corresponding to the most active layer 2 letter node is strongly activated, while all others rapidly decrease to their resting activations. The selection of the most active letter node by layer three constitutes the output of the model, but an additional

notional 'output mechanism' is interposed before the final output. This represents further output processing outside the formal domain of the model, and is included in the implementation to enable the incorporation of double letters via the geminate mechanism.

The 'decision' made by layer three is fed back to layer two via strong inhibitory one-to-one connections, so that the 'winning' layer two node is inhibited following its selection by layer three. Inhibition causes the node to take on a negative activation, from which it recovers towards its resting activation of zero at a rate determined, in the absence of excitatory input, by a parameter of the model. Since negative activations do not propagate, the inhibited node is effectively removed from the competition until it recovers sufficiently to regain a positive activation.

The geminate production mechanism itself is implemented as a further node at layer two, activated via connections from layer one in the same manner as the other layer two nodes. Whereas the letter nodes at layer two compete for output via the competitive filter of layer three, the geminate node does not compete with any other node. Instead it is held to 'trigger' when its activation exceeds a threshold, which is a free parameter of the model. The operation of the node can still be thought of in much the same way as that of the letter nodes, in that the geminate node effectively competes with its threshold rather than with other nodes.

The remainder of this section gives detailed descriptions of each element of the model. Many of the equations specified here are based on the corresponding equations of Houghton's (1990) implementation of CQ. The intention was to adapt a proven CQ architecture for this initial model rather than starting from scratch with a new model.

4.3.2 Timing Signal

Time in the model is represented by discrete time-steps. Presentation and recall of a spelling occurs at the rate of one letter per time-step, and the equations governing all operations except the competition within layer three are applied once per time-step. (The equations governing the interactions within layer three are applied iteratively several times per time-step in order to allow the competitive process to run its course

for each letter position, as discussed below). Prior to the first time-step of spelling production all node activations are set to their resting level of 0. Recall of a word is thus not affected by previous states of the model.

The spelling process is driven by the dynamic activation pattern of the I and E node at layer 1. In their original form (Houghton, 1990), the activation profiles of the I and E nodes were related by the equation

$$A_E(t) = 1 - A_I(t)$$

Equation 4.1

where $A_I(t)$ and $A_E(t)$ respectively denote the activation of the I and E nodes at time *t*, and

$$A_I = \delta_1^{\,t}$$

Equation 4.2

where *t* denotes the current time-step, and δ_1 is a parameter which sets the rate of change of the layer one timing signal. ($t \geq 0$ and $0 < \delta_1 < 1$). This results in activation profiles of the form shown in Figure 4.2(a).
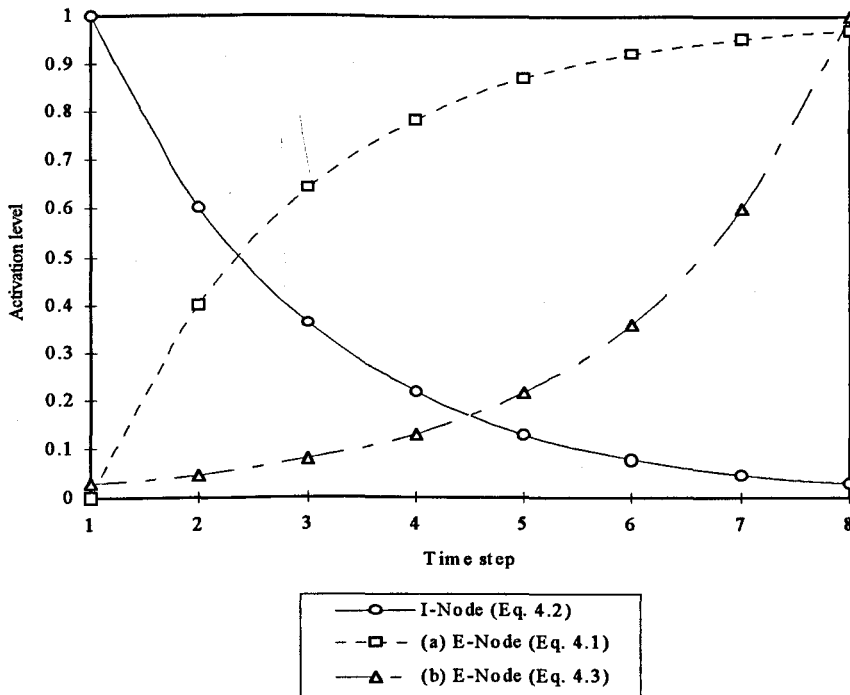
Figure 4.2. *I and E node activation profiles against time during production of a seven letter word. (a) using equations 4.1 and 4.2, (b) using equations 4.2 and 4.3.*

There are, however, many other possible forms which these activation profiles might take. One interesting alternative form has symmetrical forms for the I and E node profiles, as shown in Figure 4.2(b). This uses the same equation, 4.2, for the I node, but the E node activation is given by:

$$A_E = \delta_1^{(l-t)}$$

<div align="right">Equation 4.3</div>

where l is the number of letters in the word. In this case the E node activation rises to a well defined peak at the end of the word, rather than approaching its maximum value asymptotically as for Equation 4.1, and its activation may be thought of as a measure of distance to the end of the word. This second form relates position to the start and end of the sequence, whereas the first form relates position to the start of the sequence only. The second form therefore provides a higher level of positional resolution, and it is this form which will be employed in the model (equations 4.2 and 4.3).

4.3.3 Letter node activations

At each time-step, the activation level of each layer two (letter) node is a function of the net input it receives from the I-E pair at layer one, the inhibitory input received from layer 3, and the previous activation of the node. The activity $A_i(t)$ of letter node i at time-step t is given by:

$$A_i(t) = \delta_2 A_i(t-1) + S.f(net_i(t)) + \eta$$

<div align="right">Equation 4.4</div>

where $net_i(t)$ is the net input to node i at time t. η represents the random noise added to letter activations to simulate the damage resulting in GBD. The noise has magnitude N, and is rectangularly distributed and symmetrical about 0. A new independent noise value is generated for each letter node, and at each time-step. S operates as a variable "gain control" parameter which modulates the effect of the net input depending on the current level of activation of the unit. It is defined by:

$$S = \begin{cases} 1 - A_i(t-1) & \text{if } net_i > 0 \\ 1 + A_i(t-1) & \text{otherwise} \end{cases}$$

<div align="right">Equation 4.5</div>

The function f is the standard sigmoid 'squashing' function, which maps the input to the node into the range [-1, 1], and is defined by:

$$f(x) = \frac{2}{e^{-x\tau} + 1} - 1 \qquad\qquad \text{Equation 4.6}$$

The parameter τ sets the slope of the sigmoid.

The geminate node also operates according to equation 4.4. However, it uses a different value for δ_2 (δ_g).

4.3.4 Net inputs

The net input to a layer two node comprises the sum of excitatory input from layer one, self-excitatory input from the node itself, inhibitory input from other layer two nodes, and inhibitory feedback from the associated layer three competitive filter node. For letter node i at any time-step:

$$net_i = A_I W_{Ii} + A_E W_{Ei} + e_2 [A2_i]^+ - b_3 [A3_i]^+ - \sum_{j=1, j \neq i}^{n} m_2 [A2_j]^+ \qquad \text{Equation 4.7}$$

where $A2_i$ and $A3_i$ are activations of node i in layers two and three respectively, A_I and A_E are the activations of the I and E node respectively in layer one, W_{Ii} and W_{Ei} are the weights from the I and E nodes to node i, e_2, m_2 and b_3 are the magnitudes of self-excitation and mutual inhibition at layer two and back-inhibition from layer three, respectively, and n is the number of nodes in layers two and three. The notation $[x]^+$ indicates that only positive activation values are propagated, i.e.,

$$[x]^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \qquad\qquad \text{Equation 4.8}$$

The net input to the geminate node is also given by equation 4.7. However, no layer three node exists to provide the suppression in the fourth term of the equation. The geminate node is assumed to 'fire' when its activation exceeds that of any letter node, and the fourth term of equation 4.7 is replaced by a constant value, Inh_g.

The activations of layer three (competitive filter) nodes are updated according to equation 4.4 (with a different decay rate δ_3 replacing δ_2). For the competitive filter, the net input to each node comprises the sum of excitation from the corresponding layer two node, inhibition from other filter nodes and self-excitation:

$$net_i = c_2[A2_i]^+ + e_3[A3_i]^+ - \sum_{j=1,j\neq i}^{n} m_3[A3_j]^+ \qquad \text{Equation 4.9}$$

where c_2 is the weight of the layer two to layer three connections, e_3 is the magnitude of the self-excitatory links to layer three nodes, and m_3 is the magnitude of mutual inhibition between layer three nodes.

Applying equation 4.4 at layer 3 results in an increase in the activation of the initially most active node, and a decrease in the activations of all other nodes. To achieve a state where the 'winner' is maximally activated and all 'losers' are at zero activation requires that the equation be iteratively applied. The number of iterations required to approximate to this desired state depends on how close the activations are to start with, but experience has shown that 20 iterations are sufficient whatever the starting state. Equation 4.4 is thus iteratively applied 20 times within each time-step to allow the competitive filter to operate reliably.

4.3.5 Learning

The learning of a new sequence involves setting the values on the weights from the timing layer (layer one) to the item layer (layer two) so that the correct sequence of item selections occurs when the I and E nodes run through their standard activation profiles at recall. Houghton (1990) defined a two-phase learning procedure for the form of the CQ architecture adopted in the current model, which enables sequences to be learned from exposure.

In the first phase, the weights are set to values that directly reflect the positioning of layer two items in the desired sequence. Thus items which are to appear near the beginning of the sequence are given a strong connection to the I node and items which are to appear near the end of the sequence are strongly associated with the E node. This

can be achieved using a Hebbian, single-shot unsupervised procedure (described below) which has the advantages of speed, simplicity and neurological plausibility. However, while this is generally sufficient for the correct recall of short sequences - up to two or three items in length - with no repeated letters, for longer or more complex sequences a second phase of learning is required. Sequence generation in CQ systems is a complex dynamic process, with the items competing at one point depending in non-linear ways on the outcomes of competitions in previous positions, and correct recall of long sequences and sequences with repeated items requires a slower, iterative process of · weight setting. This entails a phase of supervised training in which the weights are incrementally adjusted to give correct recall. The two phases operate as follows:

Initial exposure

Before learning commences all node activations are set to 0, and the I-node is given its maximum activation. Thereafter the I-node activation decays at each time step according to equation 4.2. During initial presentation the E-node remains at its rest activation of zero. The letters of the word to be learned are presented to the network sequentially, one per time step. Presentation of a letter is achieved by setting the activation of the associated letter node at layer two to 1.0. The sequence is pre-processed to identify any double letters, which are presented in a single step by setting both the appropriate letter node and the geminate node activations to 1.0. The activations of all layer two nodes not set to 1.0 in this manner decay on each time-step according to the equation:

$$A_i(t) = \delta_2 A_i(t\text{-}1) \qquad\qquad \text{Equation 4.10}$$

where δ_2 is the same constant decay parameter as that used in recall, equation 4.4.

The weights on the layer one to layer two connections are updated at each time-step according to a Hebbian (co-occurrence) learning rule: Weight W_{ij} from layer one node i to layer two node j is set according to the equation:

$$W_{ij} = \varepsilon A_i A_j \qquad\qquad \text{Equation 4.11}$$

where A_i is the activation of layer one node i, A_j is the activation of layer two letter node j, and ε is a constant which controls the strength of the weights developed in this phase.

Following presentation of the final letter of the word, the I node to letter weights have values which reflect the activation of the I node at the time they were presented - the earlier the presentation, the larger the weight, with an exponential fall-off in weight strength with sequential position. The activation levels of the letter nodes have the opposite pattern, exponentially increasing in activation the more recently they were presented. At this point, the E node is fully activated and a further weight update takes place according to equation 4.11. This copies the letter node activation levels into the E-node weights, thus giving stronger weights to later letters.

This concludes the initial exposure phase. The weights thus set are able to support recall of short words of two or three letters. Longer sequences and those containing repeats require further subtle adjustment of weights by the second 'practice' phase of learning.

Practice

During the practice phase the word is recalled and the correct version is presented concurrently so that at each time-step the system has access to the correct response. If an incorrect letter is produced, the weights from the I and E-nodes to the corresponding layer two node are reduced slightly, and the weights to the node corresponding to the correct response are slightly increased. Note that random noise is not added to the node activations during practice (η in equation 4.4 is set to 0). Houghton (1990) showed this procedure to be equivalent to a gradient descent.

The following equations are applied separately to the weights from both the I node and the E node to the letter node which should have 'won' (equation 4.12) and the node which actually did 'win' (equation 4.13).

To reinforce the correct response:

$$\Delta W_{correct} = \rho(W_{max} - W_{correct})A_{IE}A_{correct}$$

Equation 4.12

And to 'punish' the incorrect response:

$$\Delta W_{actual} = -\rho W_{actual}A_{IE}A_{actual}$$

Equation 4.13

In equations 4.12 and 4.13 the subscripts *correct* and *actual* refer to the letter node which should have won the output competition and that which actually did respectively, and ρ is a small 'learning rate' parameter. The equations are applied separately to the weights from both I and E nodes, and A_{IE} represents the activation of the I or E node as appropriate. The term W_{max} in equation 4.12 imposes a ceiling on weight values.

Equivalent equations govern learning of the weights from the I and E nodes to the geminate node. If the node is not triggered when it should be, the weights W_G to the geminate node from the I and E nodes are both increased according to:

$$\Delta W_G = \rho(W_{max} - W_G)A_{IE}A_G$$

Equation 4.14

And if the node is triggered when it should not be, the weights to the I and E nodes are reduced:

$$\Delta W_G = -\rho W_G A_{IE}A_G$$

Equation 4.15

The process is repeated, with equations 4.12 to 4.15 applied iteratively, until correct recall is achieved. An important consideration when a training procedure such as this is applied in a CQ framework is that an initial error in sequencing may cause further errors simply by virtue of the incorrect item being refractory. Thus if a letter is erroneously produced too early, then in the position where it should occur it will probably still be in its refractory state, and a further error may result. This behaviour can result in many unnecessary weight changes being made by the above training procedure, which disrupts the training process. Accordingly, training is assumed to be subject to on-line repair - when an incorrect letter is produced equations 4.12 or 4.13 are applied, but the

target letter rather than the erroneous winner is subsequently inhibited. The geminate node is treated similarly - it is only inhibited following its target position.

Overlearning

The practice algorithm as specified above, and by Houghton (1990), does not allow continued practice to develop ever more robust representations, as might be expected intuitively. The competitive process of letter output is an 'all or nothing' affair, and once the correct letter wins the competition at a certain point in the sequence no further learning takes place on that letter, although it may win by a very small margin and thus be particularly vulnerable to noise.

One way to improve the robustness of learned representations would be to allow noise to affect letter activations during learning as well as recall (although at a lower level). Small winning margins would be susceptible to noise, and with enough runs of the model eventually the stochastic application of the learning procedure to affected letters would ensure a winning margin for each letter at least in excess of the noise magnitude. It is possible to regulate the robustness of the representations developed during practice in a more controlled and direct manner, however. In the model, a minimum 'winning margin', ϕ, is specified, and this is taken into account when determining which letter node is to win the output competition at each time step during practice. If the correct letter wins, but by less than the minimum margin ϕ, its closest rival is deemed to have won in its place. With this modification the practice procedure detailed above automatically continues until every letter wins by at least the specified margin. By varying the winning margin parameter the minimum robustness of a spelling's representation may be controlled.

4.3.6 Stopping

As discussed in section 4.2, spelling is to halt on the production of a 'stop symbol' at the output of the model. This symbol should be treated as a letter by the model, so it is convenient for the present purposes to label the last letter of any word as the stop symbol. This strategy is no doubt over simplified, but it will suffice for the current purposes.

4.4 Simulations and discussion

Having described the model in detail, this section examines its performance in the face of disruption by the addition of noise to letter node activation levels. To recap, the aim of these simulations is to see to what extent the gross features of GBD spelling may be simulated by the most straightforward type of disruption to the model - the addition of unstructured noise to letter node activations.

4.4.1 Simulation procedure

Test materials

Since the model makes no distinction between different letters the actual identity of the letters in a target word cannot affect the model's performance. The structure of the target word with respect to repeated letters and overall number of letters is likely to affect performance, however. The equivalence of all letter nodes means that, for example, the words 'BRONZE' and 'CHROME' are treated identically by the model - both are six-letter words with no repeated letters. Similarly 'PILLOW' and 'HATTER' (six letter words with the third letter doubled and no other repeats) are equivalently represented. Since letter identity in itself is not important it is convenient to use pseudo-words when testing the model's performance in order that the structure of the stimuli is well controlled and uniform. In all the following simulations artificial pseudo-words are used as stimuli, and the same stimulus is used on all the runs in a particular trial. The basic form of stimulus word is the letter sequence "ABC...", extended to the appropriate number of letters. Unless otherwise stated all simulations use 4000 runs of the model at each word length using letter strings with no repeated letters.

Error classification

The errors made by GBD patients AS (Jonsdottir et al. 1996) and LB (Caramazza and Miceli, 1990) have been analysed in some detail. The bulk of this analysis has taken place on words of six letters. To better compare the performance of the model with that of these patients the errors made by the model on six-letter words were analysed using the same classification system as the patients, a scheme introduced by Caramazza and

Miceli (1990). The errors made by the model are accordingly classified under five categories:

Insertions (e.g. CINEMA → CINREMA)

Deletions (e.g. CINEMA → CINMA)

Exchanges (e.g. CINEMA → CENIMA)

Shifts (e.g. CINEMA → CNEIMA)

Substitutions (e.g. CINEMA → CINOMA)

The simulations reported below involve very large numbers of runs of the model. This makes manual analysis of errors impractical, so software tools were developed to automatically classify errors. However it is difficult to consistently analyse responses which contain more than one error using automatic algorithms. The analysis was therefore limited to those responses which contain (and thus appear to have resulted from) only a single error of one of the types listed above. Some of the erroneous responses classified in the GBD studies contained more than one individual error but the majority did not. The difference in classification procedure for experimental and modelled results should therefore be small, and it was considered safe to proceed on the assumption that this will not lead to a major discrepancy between the analyses.

Following Caramazza and Miceli (1990) a scoring procedure allocates points to particular serial positions when errors occur. In the case of substitution errors it is obvious at which letter position the error occurred, and this position receives one point. Deletion errors receive one point in the position from which the deleted letter originated. Insertion errors score half a point each in the positions before and after the inserted letter in the target word. Shift and exchange errors result in two incorrect letter positions, each of which scores one point.

Parameters

The model contains a number of free parameters. Unless otherwise stated the simulations reported below all use the same parameter values, which were selected to

give correct operation of the model in the absence of noise. The degree to which the qualitative performance of the model is dependant on the precise settings of these parameters is of course an important question. This will be addressed through a systematic investigation reported in study 6 below. The actual parameter values used in the simulations are listed in Table 4.1.

Parameter	Symbol	Value
I node and E node decay rate	δ_1	0.6
Passive decay rate for layer 2 nodes	δ_2	0.6
Passive decay rate for layer 3 nodes	δ_3	0.8
Self-excitation level for layer 2 nodes	e_2,	0.1
Mutual inhibition level between layer 2 nodes	m_2	0.1
Self-excitation level for layer 3 nodes	e_3	0.1
Mutual inhibition level between layer 3 nodes	m_3	0.4
One-to-one connection strength from layer 2 to layer 3	c_2	0.4
'Back inhibition' from layer 3 to layer 2	b_3	4.0
Learning rate for initial exposure	ε	0.6
Learning rate for practice phase	ρ	0.1
Over-learning margin	ϕ	0.025
Activation function sigmoid gradient	τ	1.2
Noise magnitude	η	±0.03
Geminate node decay rate	δ_g	0.7
Geminate node inhibition level	Inh_g	1.5

Table 4.1. *Parameter values used in all simulations unless otherwise stated.*

Study 1. Basic performance - Effect of word length.

The most basic aspect of the model's performance is the overall proportion of words correctly spelled in the presence of noise. Figure 4.3 shows the proportion of correctly spelled words of different lengths with four different levels of random noise added to letter node activations during recall.
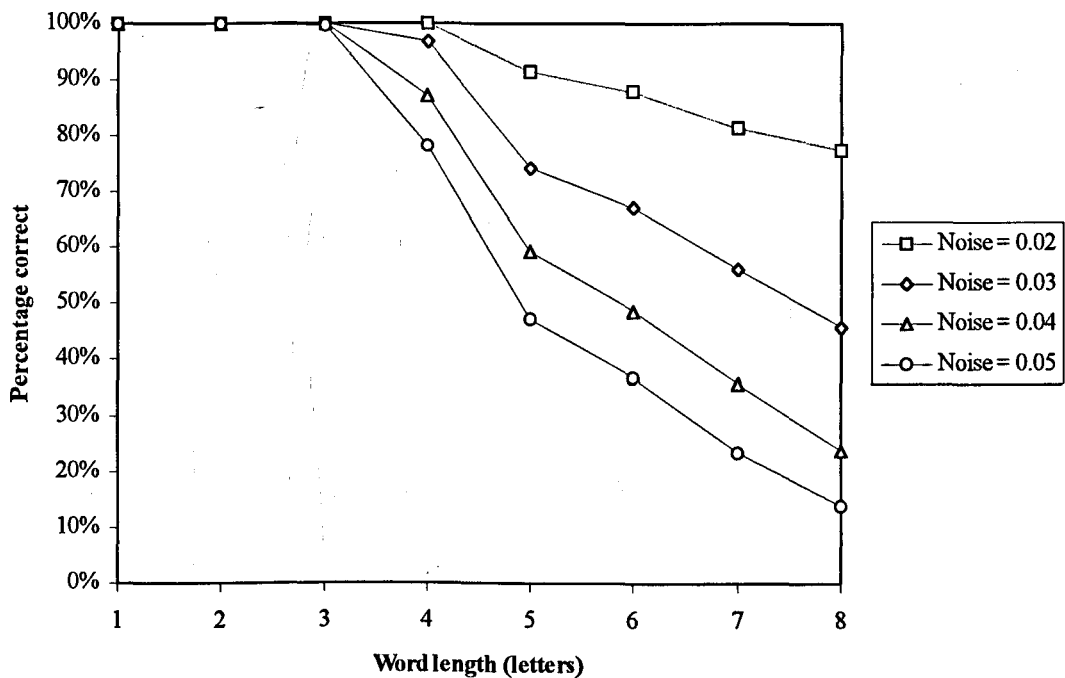
Figure 4.3. *Performance of the model plotted against word length for four different levels of random noise. The y axis shows the total percentage of words which were correctly spelled by the model over 4000 trials.*

The model shows the same qualitative pattern of performance as the patients: longer words are more vulnerable to disruption. Two factors contribute to this pattern: Firstly, since longer words contain more letters they involve more 'choice' points where the selection of the most active letter node is subject to disruption by noise. The greater number of possibilities for error leads to a lower probability of overall correct recall. Secondly, since all the serial positions of a word are represented by the same dynamic range of activity in the I and E nodes regardless of word length, longer words require that more serial positions must be discriminated within the same activation range and therefore that activation levels of consecutive letters are closer together. Closer

activations are more easily confused in the face of noise and therefore longer words have intrinsically less robust sequential representations. For a given length of word, increasing the level of noise increases the incidence of errors, as would be expected.

Figure 4.4 compares the curves for noise magnitudes of 0.02 and 0.03 in Figure 4.3 with the corresponding performance curves of three GBD patients, patient LB of Caramazza and Miceli (1990), patient AS of Jonsdottir et al. (1996) and patient JH of Kay and Hanley (1994).
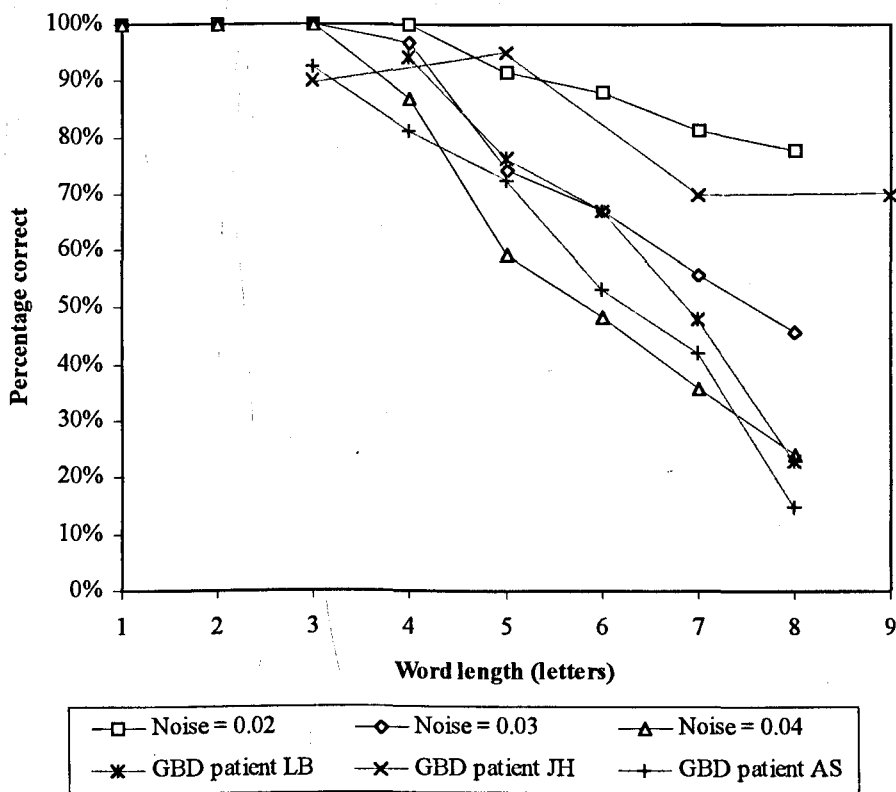
Figure 4.4. *The performance of the model compared with that of GBD patients LB (Caramazza and Miceli, 1990), JH (Kay and Hanley, 1994) and AS (Jonsdottir, Shallice and Wise, 1996).*

The 0.02 line gives a reasonable fit to the performance of JH. The 0.03 line fits the performance of LB well for words of up to seven letters, but the last point, for 8-letter words, shows a large mismatch. However, the fit for word lengths of 4, 5, 6 and 7 letters is good enough that the model's general match to these patients is encouraging. Patient AS lies generally between the 0.03 and 0.04 lines, but is not fitted by either so

well as the 0.03 line fits LB. Since the 0.03 line fits LB's performance well, and is almost as good a fit to AS as the 0.04 line, the noise magnitude was fixed at 0.03 for the remainder of the simulations.

Study 2. Serial position curves.

A marked increase in the likelihood of an error in medial letter positions is a robust characteristic of both normal spellers and GBD patients. Using the scoring scheme of Caramazza and Miceli (1990) it is possible to score each serial position for the likelihood of an error occurring. Figure 4.5 shows the resulting serial position curve for single-error responses on six-letter words with no repeats compared with the equivalent curves for patients LB and AS.
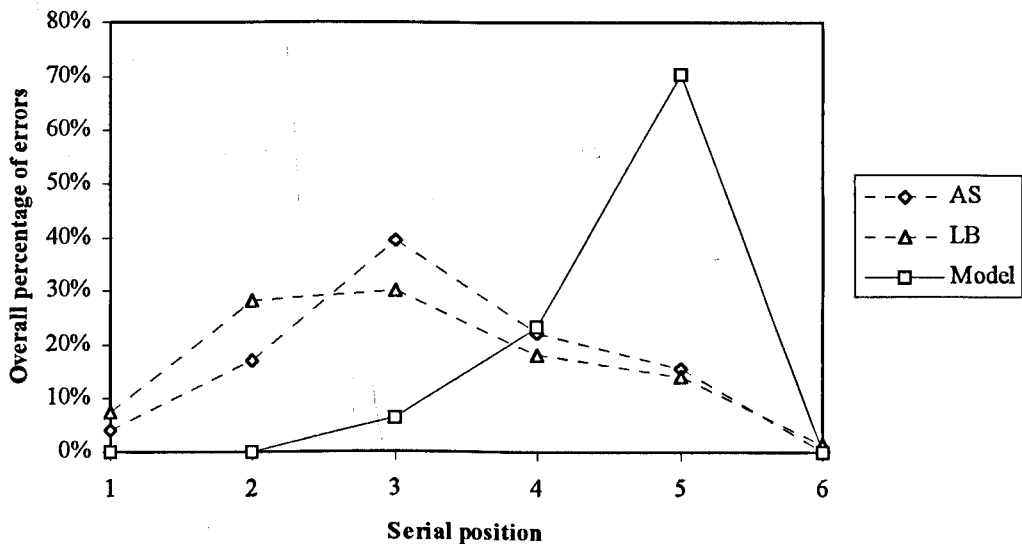
Figure 4.5. *The incidence of single-error responses made by the model at each serial position in six-letter words, given in terms of the error points awarded by the scoring procedure of Caramazza and Miceli, compared with the equivalent curves for GBD patients LB and AS.*

The model shows the correct general pattern; that is, the error rate is lowest in the initial and final positions. However the peak error rate is in letter position 5, considerably closer to the end of the word compared with position 3 for AS and LB.

The low error rates in initial and final positions is due to the dynamic behaviour of the I-E activating signal during spelling production. All the information necessary for the generation of a sequence is contained in the weights from the I and E nodes to the letter nodes, and each position in the recalled sequence is encoded by a unique combination of I and E activations. However, the distance between the I/E combinations representing consecutive positions does not remain constant across the sequence. At the beginning of the sequence, the I node activation falls very quickly, and consecutive positions differ greatly on this axis, while the E node activation is climbing only slowly. · Towards the end of the sequence, the opposite state of affairs holds, with rapidly climbing E node activation giving large differences between consecutive positions while the I node's rate of increase has slowed. In the middle of the sequence both nodes' activations are changing fairly slowly. Taken as a whole, consecutive positions towards the start and end of the sequence thus differ more in their I/E encodings than those towards the middle.

Figure 4.6 breaks down the serial error curve of Figure 4.5 to show the incidence of the five basic error types - insertions, deletions, exchanges, shifts and substitutions - at each serial position.
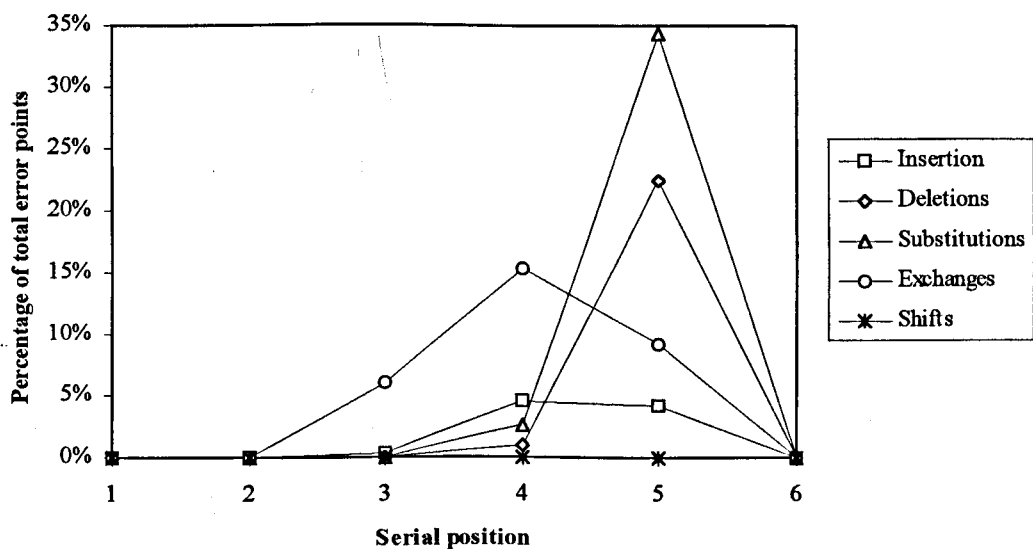
Figure 4.6. *Serial error curves produced by the CQS model on six-letter words with no repeated letters, plotted separately for each of the five basic error types.*

Each individual error type shows low error rates in initial and final positions, as do patients LB and AS. Exchange, insertion and shift errors peak relatively medially in good agreement with GBD patients. However, deletion and substitution errors peak towards the end of the word, and are clearly a major factor in the skewed overall serial error curve of Figure 4.5. In this connection it is very interesting to note that every one of the 401 substitution errors which occurred in position five during the 4000 runs of this simulation consisted of the re-occurrence of the first letter of the target word. Such a stereotyped error pattern is very different to the varied errors produced by patients, and is suggestive of a problem with the error mechanisms of the model at this level of analysis.

Study 3. Proportions of individual error types.

Figure 4.7 shows the relative proportions of the five basic error types made by the model on 6-letter words.
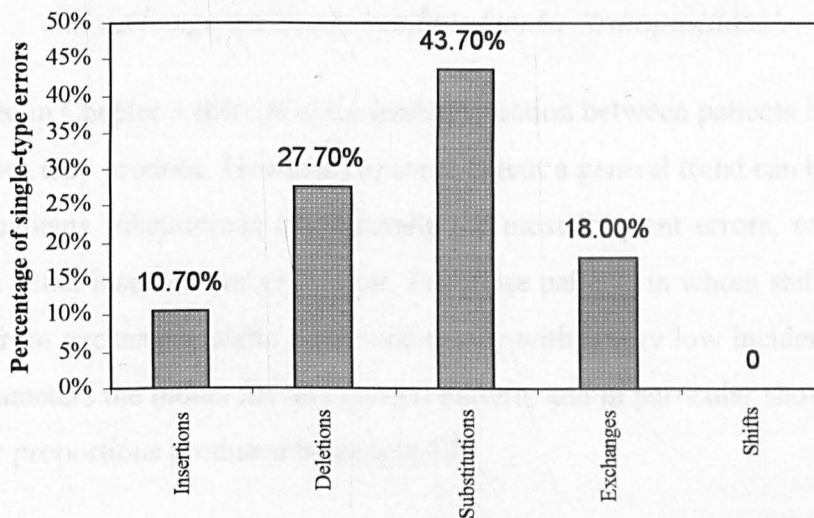
Figure 4.7. *Relative occurrences of the five basic error types analysed for the model on six-letter words.*

Figure 4.8 compares these proportions with those produced by four of the most studied GBD patients on six letter words. Shift and exchange errors are combined as one category, transpositions, since this classification is adopted for some patients.
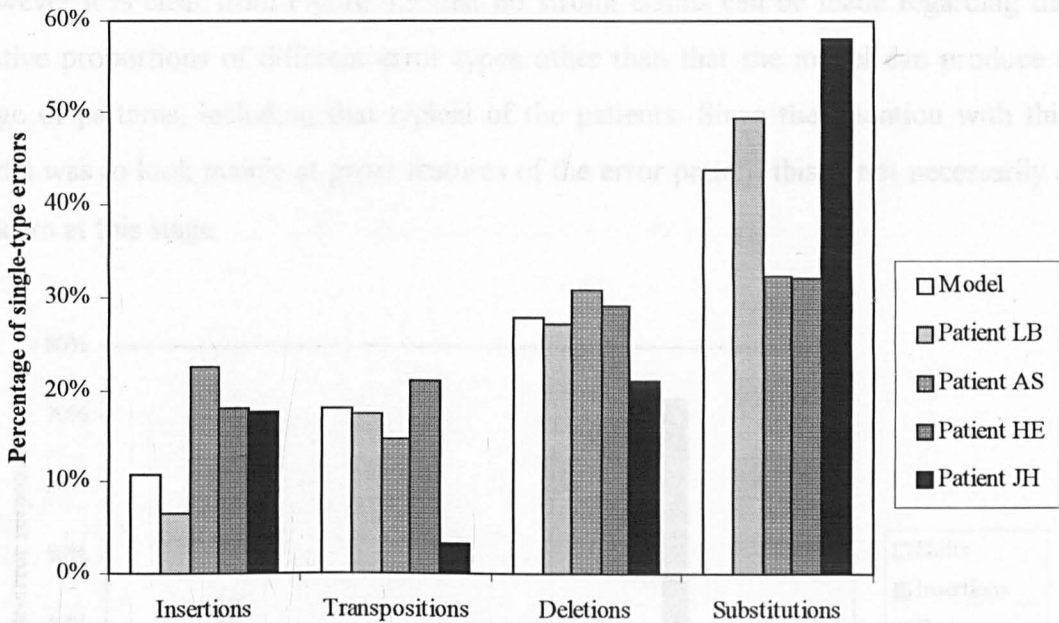
Figure 4.8. *Relative proportions of five major error types for the model compared with four GBD patients LB, AS, JH and patient HE of McCloskey et al. (1994). Shift and exchange errors are combined under 'transpositions'.*

As discussed in Chapter 3 there is considerable variation between patients in the pattern of error types they produce. However, to some extent a general trend can be discerned: For these patients substitutions are generally the most frequent errors, with deletions higher than either insertions or exchanges. For those patients in whom shifts have been separated from exchanges, shifts occur universally with a very low incidence. With its default parameters the model fits this general pattern, and in particular shows a close fit to the error proportions produced by patient LB.

This good fit must be treated with caution, however. The model's performance in this area is greatly affected by parameter variation (this issue will be explored in study 6), and the default parameter settings were chosen with an eye to demonstrating that such a fit is possible. Figure 4.9 compares the error proportions of Figure 4.7 with equivalent data from four simulations with slightly different parameters. The pattern is volatile with respect to the parameters 'layer two decay' and 'sigmoid slope' - study 6 will demonstrate that this is the case for most other parameters also. Varying the level of noise, on the other hand, has only a minor effect on the relative ranking of error types.

However it is clear from Figure 4.9 that no strong claims can be made regarding the relative proportions of different error types other than that the model can produce a range of patterns, including that typical of the patients. Since the intention with this model was to look mainly at gross features of the error pattern this is not necessarily a concern at this stage.
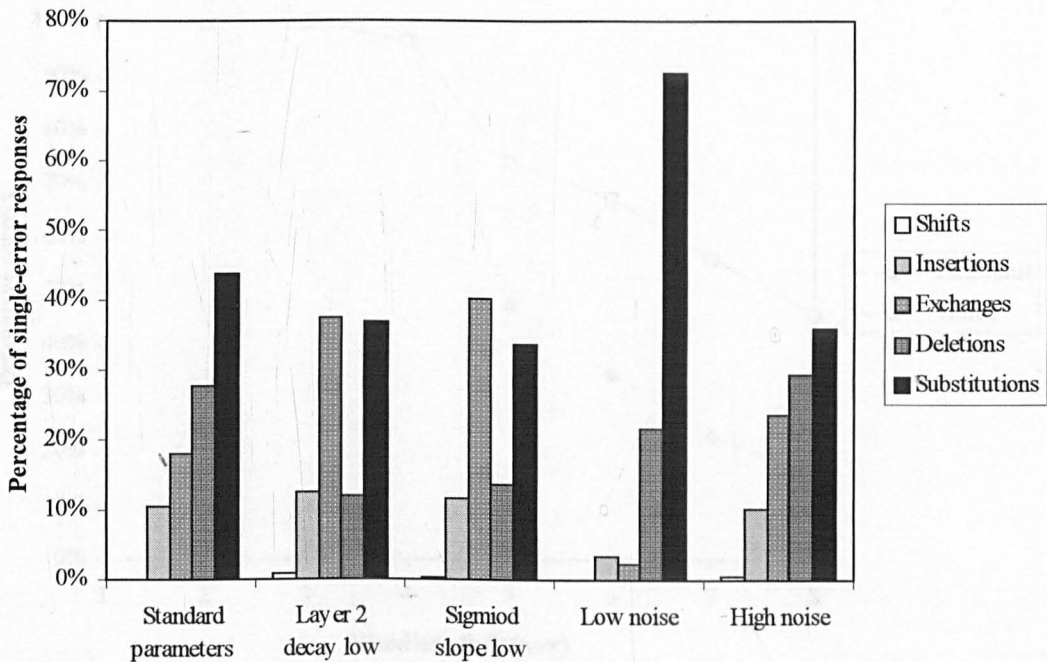
Figure 4.9. *The relative proportions of different error types produced by the model on 6-letter words using the standard parameter values compared with those produced using slightly different sets of parameter values. The 'standard parameters' condition uses the parameters of Table 4.1. In the other conditions one parameter is varied at a time. 'Low' parameter settings are 80% of those of Table 4.1, 'high' settings are 120% of the Table 4.1 value.*

Study 4. Word familiarity.

It is characteristic of GBD patients that they show a similar qualitative pattern of errors on both familiar and novel words, although their quantitative performance is somewhat worse on novel words. As discussed earlier, it was hoped that the distinction between familiar and novel words could be approached by manipulating the 'overlearn'

parameter on the model. It was suggested in section 4.3 that novel words ('nonwords') could be modelled by removing this winning margin requirement, and thus stopping practice as soon as the model is just able to produce the target word. By comparison with normal operation such words would be less well learned. Figure 4.10 compares the performance of the model on words and nonwords modelled in this way.
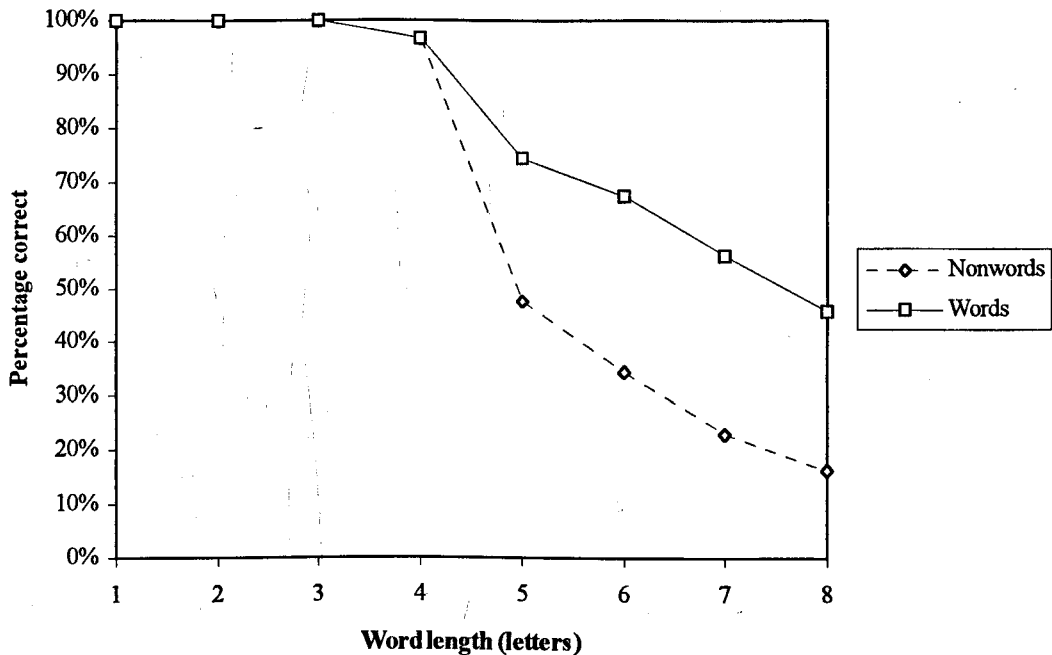
Figure 4.10. *Performance of the model on words compared with nonwords, plotted against word length. Nonwords were modelled by setting the 'overlearn' parameter value to 0.*

Clearly the model performs less well in general on nonwords than on words of the same length. This is not true for shorter words, however, of lengths two, three and four letters. The reason for this is that the model requires no practice at all for words up to three letters in length, and only very minimal practice for four letter words. The initial exposure phase of learning operates so well for words of these lengths that they start out nearly as robust as required by the 'overlearning' margin, and the absence of this margin thus makes very little difference to the robustness of their representation.

The approach can be assessed in more detail by looking at the proportions of different error types made on words and nonwords modelled in this way. GBD patients typically

produce very similar error patterns regardless of lexical status - LB, for example, produces nearly indistinguishable proportions of different error types on words and nonwords. Figure 4.11 compares the two for the model.
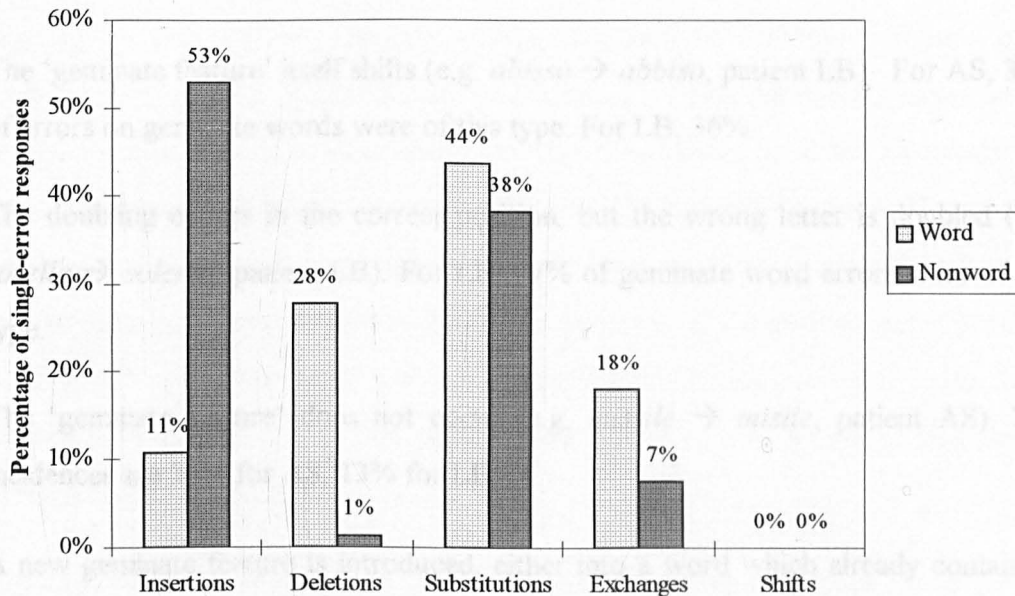
Figure 4.11. *Relative proportions of different error types produced by the model as a function of lexical status. Nonwords are modelled by setting the 'overlearn' parameter value to 0.*

The manipulation clearly makes a large difference, especially to the relative incidences of insertion and deletion errors. Figure 4.9 showed that the relative incidences of errors is sensitive to variations in the model, and since the practice phase of learning is a qualitatively different type of process from the initial exposure stage, it is perhaps not surprising that the representations developed by the two phases should be different enough to produce different error patterns. While the general approach of modelling novel words as those with less well learned representations seems attractive, there are clearly difficulties with the way this is implemented in the model.

Study 5. Double letters

GBD patients show a characteristic pattern of errors on double letters, which suggests that the property of doubling can dissociate from the letter being doubled. Patients' errors on double letters fall into several classes:

1. The 'geminate feature' itself shifts (e.g. *abisso* → *abbiso*, patient LB). For AS, 30% of errors on geminate words were of this type. For LB, 36%.

2. The doubling occurs in the correct position, but the wrong letter is doubled (e.g. *sorella* → *solerra*, patient LB). For LB, 27% of geminate word errors were of this type.

3. The 'geminate feature' does not occur (e.g. *missile* → *misile*, patient AS). The incidences are 30% for AS, 13% for LB.

4. A new geminate feature is introduced, either into a word which already contains a doubling (e.g. *pepper* → *peepper*, patient AS; For AS, 4% of errors on geminate words were of this type, for AS, 18%) or, very rarely, into a word with no doubling (e.g. *tavolo* → *tavvolo*, patient LB; For LB this type of error only occurred on 0.02% of geminate words).

The model produces errors of types 1 and 2, but not 3 or 4. On 4000 seven-letter words with a single medial double letter (in positions 3 and 4), using the standard parameters, the model made 20 errors of type 1 (1.7% of all errors) and 86 errors of type 2 (7.3% of all errors). These were the only errors which concerned the double letter alone. Clearly these incidences are far lower than those for GBD patients. However, this model was intended only to qualitatively model detailed aspects of the error pattern such as this. Given the inability of the model to produce double letters without the intervention of the geminate mechanism, even these low levels of errors are enough to demonstrate that the qualitative behaviour of the geminate system is correct, in as far as it gives rise to errors of types 1 and 2.

Study 6. Robustness of model in the face of parameter variation

Formalisation of complex models of psychological processes inevitably introduces free parameters which scale the size of various processes and interactions. The need to fix parameters to produce concrete simulations can lead to accusations of "curve fitting". Depending on the chosen parameter setting, the same model may produce quite different patterns of data. In general, the best strategy is to set parameters to generate a particular level of general performance, and then keep them fixed for all subsequent detailed explorations. This is the strategy pursued in the simulations above. However, in assessing a model it is still important to analyse how sensitive its various empirically desirable features are with respect to parameter variation. If we distinguish between the "robust" and "fine" behaviour of a model, the latter being highly parameter independent and the former not, it is generally on the basis of the robust behaviour that strong theoretical claims can be made.

The model contains a large number of free parameters which have been set to give as good a performance as reasonably possible when compared with GBD patients' spelling. This was carried out by starting from values taken from Houghton (1990), values which were known to enable correct learning and recall, and manually adjusting these until a set was found which gave good performance. A more defensible method would be to systematically scan the entire parameter space of the model, in small increments, to both find the best possible fit and to determine the area of parameter space for which the desired qualitative behaviour of the model holds good. However, with such a large number of independent parameters such a procedure would be highly impractical.

Nonetheless, it is most important - as it is with any parameterised model - to address the issue of the extent to which the results reported above have been influenced by the choice of default parameters. Although a full systematic search of parameter space is impractical, a more limited study in which each parameter is varied in turn is more tractable and will still provide some reassurance that the effects reported have not been obtained only by careful setting of parameters. With this aim, Table 4.2 shows the effect of parameter variance on the main qualitative effects reported in this section. Each row

represents 1000 runs of the model on six-letter words using the parameter settings of Table 4.1, except for the indicated parameter which is varied up and down by 20% from the default value, and the noise setting, which is adjusted to give a performance of between 65% and 67% correct, approximately the level of performance with the default parameters (apart from the rows in which noise is the parameter under study, of course). This is in order that parameter sets are compared on a level footing with respect to overall performance. The parameters varied are those which have the largest effect on performance. The meaning of the columns is as follows:

Value with respect to normal setting: Each parameter shown is varied up and down by 20% from the default value of Table 4.1.

Spelling is possible: '*' indicates that the model was able to learn and recall 6-letter words in the absence of noise.

Word length effect present: '*' indicates that performance declines for longer words.

Serial position effect present: '*' indicates that the lowest error incidence occurs in the initial and final letter positions.

Relative proportions of error types meet criterion: '*' indicates that the pattern of relative proportions of different error types is broadly stable and comparable to that found for the default parameters in study 4.3 - that is, substitutions are the most frequent errors, with deletions higher than either insertions or exchanges, and shifts occur only rarely. (In all cases, the incidence of shifts never exceeded 2% regardless of the other error incidences).

In one case, operation of the model proved impossible. In all other cases the model's performance is robust for word length and serial position effects. The relative proportions of error types are highly volatile with respect to parameter variation, however, as has already been indicated. The only parameter studied here which does not materially affect these proportions is the noise magnitude.

Parameter	Value with respect to normal setting	Spelling is possible	Word length effect present	Serial position effect present	Relative proportions of error types meet criterion
Layer 2 self-excitation	120%	*	*	*	-
e_2	80%	*	*	*	*
Layer 2 mutual inhibition	120%	*	*	*	*
m_2	80%	*	*	*	-
Layer 2 decay	120%	*	*	*	-
δ_2	80%	*	*	*	-
Excitation from layer 2 → layer 3	120%	*	*	*	-
c_2	80%	*	*	*	-
Back inhibition from layer 3 → layer 2	120%	*	*	*	-
b_3	80%	*	*	*	-
I node decay rate	120%	-	-	-	-
δ_1	80%	*	*	*	-
Sigmoid slope	120%	*	*	*	-
τ	80%	*	*	*	-
Over-learning margin	120%	*	*	*	-
ϕ	80%	*	*	*	-
Noise magnitude	120%	*	*	*	*
η	80%	*	*	*	*

Table 4.2. *Effect of independently varying key parameters of the model. See text for details.*

4.5 Summary

The aims for this model were to provide a qualitative account for four gross or 'core' aspects of GBD:

1. Word length effects.

2. Serial error curves.

3. The occurrence of insertion, deletion, shift, exchange and substitution errors.

4. The types of error made by GBD patients on double letters.

The model has provided a qualitative, and in some cases also quantitative, fit to the data in all these areas, and has thus succeeded in fulfilling these aims. In particular, it has demonstrated that the effects of word length, serial position and double letters in GBD are explicable with a CQ account.

As is to be expected of an initial attempt at applying a model in a new domain, the model has a number of shortcomings, mostly with respect to more detailed aspects of its behaviour. These are as follows:

1. **Word length effects.** The performance vs. word length curves fit the patients' performance quite well, but not perfectly. The curves have a slight bow which appears to be in opposite directions for the patients and model, and long words in particular do not fit well.

2. **Serial position effects.** The model shows a bowed serial error curve, but errors peak much later in the word than they do for GBD patients. This is particularly the case for substitution and deletion errors.

3. **Error types.** The model shows a high incidence of substitution errors involving the re-activation of the first letter in the last-but-one position. This is at variance with GBD performance.

4. **Error proportions.** The model is able to show relative proportions of different error types in good accordance with GBD data. However, this pattern is highly labile and is adversely affected by most parameter changes.

5. **Effect of lexical status.** The manipulation used to model the word/nonword difference - variation of the 'overlearning' margin parameter - was only partially successful. While the correct qualitative behaviour was produced with regard to overall performance, this was only the case for longer words where a significant amount of supervised practice is necessary for correct recall. Moreover, the relative error proportions are affected by this manipulation, which is not the case for word/nonword differences in GBD patients. An alternative manipulation which could have been used is the alteration of the noise magnitude. This both smoothly improves or degrades overall performance over the entire word length scale, and leaves the relative proportions of different error types unaffected. Conceptually, the two manipulations should be equivalent - increasing the amount of learning would be expected simply to make the representation of a word more robust in the face of noise. However, in practice the supervised phase of learning appears to qualitatively change the representation rather than smoothly increasing its robustness.

6. **Double letters.** Two striking effects - the movement of the 'geminate feature' and errors affecting the identity of the doubled letter but not its position - are both shown by the model, bearing out the use of a separate and error prone geminate mechanism for double letter production. However, the model never deletes or inserts geminate features, in contrast to GBD patients.

7. **Parameters.** The large number of free parameters in this model makes it difficult both to set up the model and to investigate the parameter space.

The good basic performance of the model justifies its use as the basis for further development. Future models will need to resolve the issues listed above if the finer aspects of GBD are to be convincingly modelled, but these will be addressed in parallel with development of the model to address the remaining major feature of GBD spelling - that of consonant - vowel status preservation in errors. The next chapter will therefore

explore the implications of representing CV status in CQ spelling models, before the modelling thread is picked up again in Chapters 6 and 7.

5

Serial constraints in Competitive Queuing[4]

The main feature to be addressed in the further development of the CQ approach to spelling is the preservation of consonant/vowel status in GBD spelling errors. Rather than address the CV constraint in an ad-hoc manner, it is sensible to consider general approaches to the implementation of serial constraints in CQ models which will be of wider utility and theoretical interest. This chapter examines the issue of serial constraints in general and identifies two distinct classes of mechanism which may be used to implement such constraints in a CQ system.

5.1 Serial constraints in sequential behaviour

Chapter 1 identified serial constraints operating in speech and verbal short-term memory, as well as GBD spelling. In the domain of speech production a fairly tight set of constraints governs which phonemes may legally be present at which stage of syllable production (Hartley and Houghton, 1996). Thus in English the syllable *'flunt'* is legal, but *'lfunt'* is not. When errors occur on individual phonemes within a word, the erroneous phoneme is virtually always one which is legal in the syllable position. The constraint on errors is particularly striking in the case of exchange errors. Thus the error

BARN DOOR → DARN BORE

where two onset phonemes have exchanged, is quite possible, whereas the exchange of an onset phoneme with a coda phoneme:

---

[4] The main theoretical point made in this chapter was first made in Glasspool and Houghton (1997).

BARN DOOR → OORARN DB

is most unlikely. In verbal short-term memory one can also identify the operation of serial constraints. Thus errors in which two items exchange are particularly likely to involve similar sounding items (Conrad, 1964), and when errors occur in the recall of grouped lists, items which change groups tend to retain their within-group position (Fuchs, 1959).

Such constraints bear a strong family resemblance to the CV preservation constraint in GBD spelling, and it would be satisfying to develop an overall account for the imposition of constraints on categories of responses at specific serial positions within the CQ paradigm. With this in mind it is worth noting some common characteristics of these apparent serial constraints:

1. Serial constraints are *abstract* in the sense that they cannot be expressed in terms of specific actions, but instead involve whole classes or categories action. Whereas the other common features of error patterns in different domains discussed in Chapter 1 are similar across domains, the categories on which serial constraints operate are of necessity domain-specific - in speech they operate on classes of entities such as phonemes and words, and in spelling on consonant and vowel letter classes.

2. Categorical constraints appear to vary in the ease with which they may be violated, largely according to the domain in question. Thus violations of syllabic position in the source and destination of phoneme movement errors in speech are extremely rare, while as many as 15% of errors in GBD patient AS violate the preservation of CV status.

How do these various characteristics fit the general behaviour of the Competitive Queuing architecture?

5.2 Serial constraints in CQ models: Biasing the queue

In an unconstrained CQ system movement errors are common, their foremost characteristic being that the shorter the distance moved, the more likely the error. This property is due to strong competition between nearby items in CQ. The existence of a

serial constraint implies that each item competes mainly with nearby members of its own category. This suggests a simple model - multiple queues appear to be in operation, one for each category of responses (Figure 5.1). On this model the existence of a CV constraint in GBD spelling would be explained by assuming that in positions where a consonant appears in the target word, only consonants compete for output during spelling, and only vowels compete for output in vowel positions.
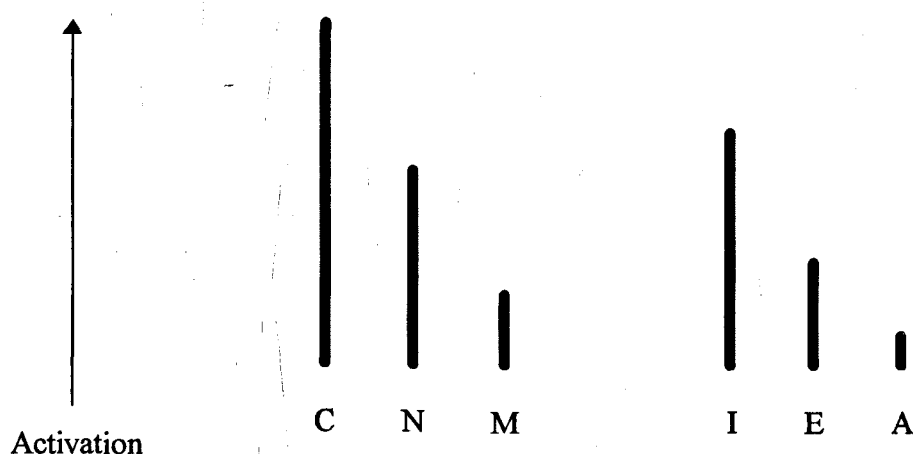
Figure 5.1. *An outline model for the operation of a serial constraint on CV structure in a CQ spelling system. Separate queues operate for consonants and vowels. Only consonants compete for consonant positions; only vowels for vowel positions.*

Errors in such a system would both preserve consonant/vowel status and would preserve the usual CQ error pattern of higher probability of movement errors the shorter the distance moved, but the pattern would operate within, but not between, categories. For example, the exchange error CINEMA → CENIMA would be a likely error. In the second letter position only vowels compete, the most active being the target letter, **I** (right hand side of Figure 5.1). If an error occurs in this position the most likely erroneous letter is **E**. The third position is a consonant position and will be unaffected by the error in the vowel queue. In the fourth position the vowel queue is again in control of output, and with the target letter **E** still partially inhibited following its erroneous selection a further error is likely, the most probable candidate being **I**.

There are two problems with this simple model. Firstly, the switch between alternative queues is absolute, which implies that there can be no possibility of errors violating category constraints. While this may approach the truth in domains such as speech production, it is certainly not true of GBD spelling or of verbal STM, where category constraints are far from absolute and errors often violate them. The second problem is the complex extra machinery required in order to maintain multiple simultaneous queues and switch between them during sequence production. However, as Glasspool and Houghton (1997) point out, it is possible to overcome both of these difficulties by · retaining a single queue but using category-specific biasing to create multiple competitions within that queue.

5.2.1 Serial constraints in a single queue

Consider again the production of the word CINEMA, but in a system with a single queue, with a small constant bias added to the activation of all consonant letters in consonant target positions and to all vowel letters in vowel target positions (Figure 5.2).

Activation   C I N E M A   C I N E M A

Bias added in Vowel   *Bias added in Consonant*
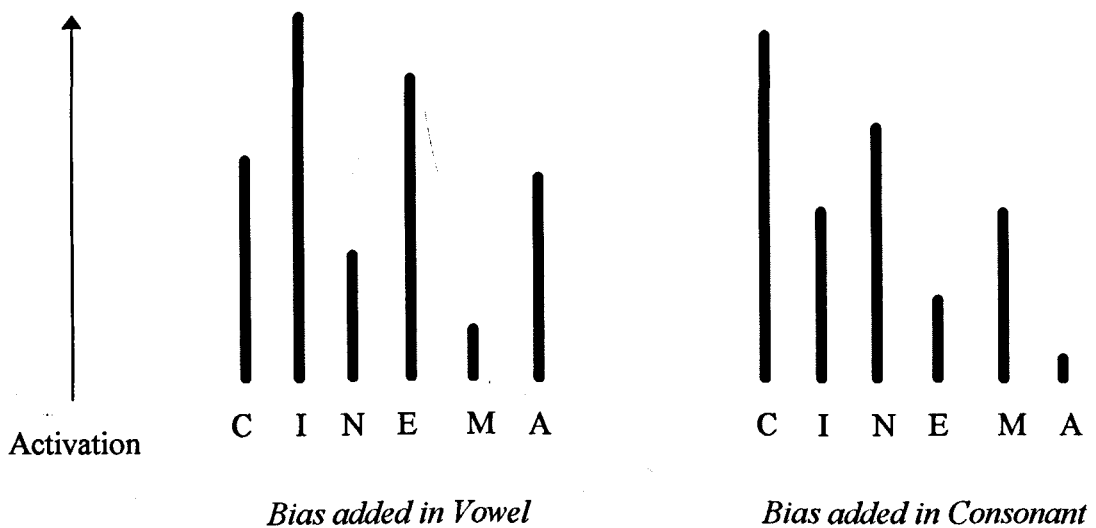
Figure 5.2. *A CV constraint in a single queue with a dynamic bias, which raises the activity of all vowel letters in vowel positions and all consonant letters in consonant positions.*

This manipulation creates multiple competitions within a single queue. The qualitative effect is similar - the main competitors in each position are other items of the same category, which have been boosted above the level of nearby items of other categories by the bias. However, the constraint is not absolute like that imposed by separate queues. The degree to which the competitions separate can be varied smoothly and continuously by varying the bias: a large bias gives a complete separation, with little or no chance of category violations. A small bias gives only a slightly higher chance of category status being preserved in errors than no bias. (There is of course no need for · the bias to be in the positive direction; the approach would work just as well if items which do not fit the category of the current position are inhibited). Only a single competitive queue is required whatever the number of categories which must be differentiated, and the machinery required to switch categories is likely to be greatly simplified by comparison with that required to switch between queues.

Where might such a bias originate? It is possible to identify two classes of biasing arrangement, which may be termed *internal* and *external* constraint systems. In an external constraint system the bias originates in a separate mechanism, which imposes serial constraints by providing a dynamic bias to certain subsets of competing items - corresponding to classes of response - at appropriate points during sequence production. In an internal constraint system the bias emerges from interactions between the internal representations of items within the CQ sequencing system itself.

5.2.2 External constraints

In the external class of constraint system the bias comes from a separate mechanism which applies a categorical 'template' to the production of sequences. The template can be thought of as embodying a dynamic grammar which is imposed on the sequences generated by the underlying CQ system (Figure 5.3).
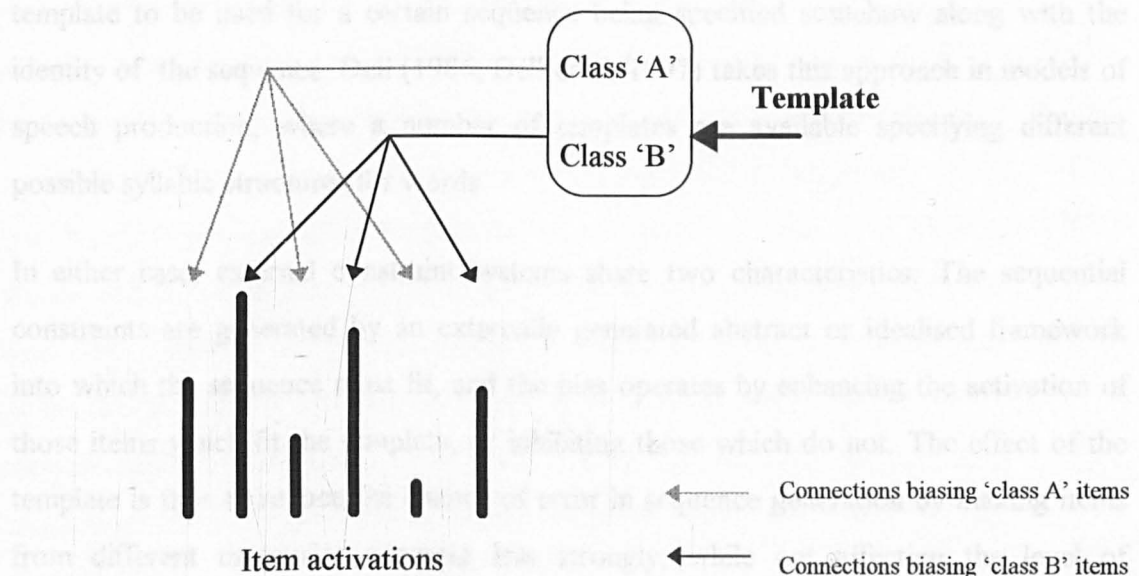
Figure 5.3. *The "external" class of constraint system employs a mechanism additional to the basic CQ sequencing system which imposes serial constraints by providing a dynamic bias to certain classes of response during sequence production.*

Models of this type have been important in psycholinguistically oriented work on language and speech production (see e.g., Dell, 1986, 1988; Dell, Burger and Svec, 1997; MacKay, 1972, 1987; Hartley and Houghton, 1996; Martin, Dell, Saffran and Schwartz, 1994; Stemberger, 1985).

The template may be highly stereotyped, as it is in Hartley and Houghton's (1996) model of speech production. A 'syllabic template' is applied to the sequencing system in this model to constrain the set of possible phonemes at each step in the production of a word according to the position within the syllable. The constraint operates by boosting the activation levels of all phonemes which are legal in the current syllabic position, making it less likely that an error will involve a phoneme from outside the 'boosted' set. The same syllabic template is applied to all phoneme sequences produced by the model. (The approach is similar to that of Vousden, 1996, as noted in Chapter 2.)

Alternatively, a set of different structural frameworks may be available, each specifying a different possible 'template' into which sequences may be fitted, the particular template to be used for a certain sequence being specified somehow along with the identity of the sequence. Dell (1986; Dell et al, 1997) takes this approach in models of speech production, where a number of templates are available specifying different possible syllabic structures for words.

In either case, external constraint systems share two characteristics: The sequential constraints are generated by an externally generated abstract or idealised framework into which the sequence must fit, and the bias operates by enhancing the activation of those items which fit the template, or inhibiting those which do not. The effect of the template is thus to reduce the chance of error in sequence generation by making items from different categories compete less strongly, while not affecting the level of competition between items in the same category.

5.2.3 Internal constraints

In an internal constraint system the bias originates in interactions between representations within the sequencing system itself. This can occur when the internal representations of similar items interfere and items receive partial activation from similar items in the same sequence. Such an effect differs from that of external constraint systems in two main ways: Firstly, no stereotyped template is imposed on the sequence - the structural template for the sequence, such as it is, is formed by interactions within the sequence itself and is not limited by any form of underlying grammar. Secondly, the constraint operates through interference between similar items, and thus in contrast with the external type of constraint it acts to increase the chance of error and thus reduce the robustness of the sequencing system.

The operation of the internal type of constraint can be best illustrated by means of an example. A number of models of verbal short-term memory have been based on the CQ approach (Burgess and Hitch, 1992, Glasspool, 1995, Burgess and Hitch, 1996), as discussed in Chapter 2 (Figure 2.5 shows the general arrangement of such models). Such models offer essentially similar explanations for the effect of phonological similarity in verbal short-term memory. As discussed in Chapter 1, It is well known that

memory for lists of words is impaired when the words are phonemically similar. On the model, these effects are due to overlap between the phonemic representations of phonemically similar words. A stimulus word at the input to the model not only activates its associated word node but also partially activates those of any phonemically similar words, which thus become weakly associated with the current state of the context signal during the learning process. During recall phonemically similar words receive additional activation from their weak association with each other's context signals, and thus compete more strongly with each other than would otherwise be the case. This increased competition results in higher rates of errors in lists containing similar words. Errors are particularly likely to involve confusions between pairs of phonologically similar words, as such pairs effectively bias each other. The bias in this case is due entirely to overlapping internal representations for similar words.

5.3 Discussion

An external constraint represents an additional source of information which should tend to improve performance by reducing the chance of confusion between items of different categories, while not affecting the chances of confusion between items within the same category. The internal type of constraint, by contrast, represents a reduction in the distinctiveness of items within categories - similar items interfere by partially activating each other. The separate template and associated mechanisms required to implement an external constraint system impose an extra overhead, the benefit of which is a reduction in errors on sequences with mixed categories. Internal constraint systems only ever increase error rates. In either case, however, the result is qualitatively similar - performance is worse on sequences which only contain items of the same category compared with sequences of items from different categories, and errors show a tendency for items of the same category to interact.

There are two points of difference, though, one pragmatic and one structural, which may help distinguish which approach is most appropriate in any particular domain. From a pragmatic point of view, the more stereotyped abstract categorical structures are in a domain the greater the payoff from the additional machinery of the external approach, since one or a few highly stereotyped templates may capture most of the

productive potential of the domain. For example, Hartley and Houghton's (1996) speech production model uses a single template to cover the majority of syllable structures in English. Domains where categorical structure is completely unconstrained, such as verbal STM, would need as many templates available as there were possible categorical structures, and so the external approach would be less reasonable in this case. From a structural point of view, the external approach involves an additional mechanism, the template, which it may be possible to detect experimentally. Sevald, Dell and Cole (1995) attempted to determine if a set of templates for different C/V structures exists in the speech production mechanism, as suggested by Dell (1986) for example. They carried out experiments in which subjects repeated sequences of nonwords, the second of which could share phonemes and/or syllabic structure with the first. They found higher speech rate for the second word if it shared syllabic structure with the first, regardless of whether or not it shared any phonemes. This is consistent with the view that separate structures carry syllabic structure information and phoneme identity information, and that it is the former which is being primed here. It is possible that this type of experimental work may be able to tease apart the different possibilities empirically.

5.4 A C/V constraint in models of spelling

In terms of the number of different abstract structures which are possible, spelling appears to be intermediate between speech and STM. There are some constraints - spellings generally don't contain all consonants or all vowels - but these are not as hard and fast as those for CV structure in speech, where certain language-dependent rules are never violated by legal words. If a relatively small number of CV structures encompass a large proportion of frequent spellings it might make economic sense to separate out the abstract CV structure as an external template system. Empirical studies such as that of Sevald et al. (1995) have not been carried out in the domain of spelling, so the question of whether an external or internal constraint system would best model the preservation of CV status in GBD errors remains open.

An interesting course to take in developing the CQ spelling model to encompass CV constraints will therefore be to explore the consequences of choice of internal or

external constraint approaches. Accordingly, the next two chapters will advance, in parallel, alternative developments of the CQS spelling model of Chapter 4. Chapter 6 will use an external CV template to specify the CV status of each letter during the generation of a spelling, while Chapter 7 will explore the possibilities opened up by the use of distributed representations internally to the model, thus allowing letters of the same CV class to interfere with one another.

6

CQX: A CQ spelling model with external constraints

Chapter 5 distinguished between internal and external sources of constraint in CQ mechanisms. In this chapter, external constraints are investigated through the addition of explicit CV status information to the spelling mechanism. The CQS architecture will be taken as a starting point, and a second aim of this chapter is thus to address the main shortcomings of that model listed in section 4.5.

6.1 Design considerations

6.1.1 Improving positional representation

Several of the shortcomings of the CQS model suggest that increasing the positional discrimination of the activating system would help. In particular, there is a tendency in CQS for letters to be re-activated towards the end of words (This contributes to both problems 2 and 3 of Section 4.5). If the positional discrimination of the sequencing system is improved to the point where it is possible for a temporally well-defined activation peak to be produced on the geminate node, this will also open the possibility of geminate insertion and deletion errors. The first change to the CQS model will thus be to increase positional resolution.

Chapter 2, section 2.7, discusses the two ways in which positional discrimination in CQ models may be increased - by increasing the dimensionality of the input (or timing) signal, or by increasing the positional resolution of the activation function. While a number of CQ models have taken the former approach to improving positional discrimination (e.g. Burgess and Hitch, 1992, 1996, Glasspool, 1994, Brown, Preece and Hulme, submitted, Vousden 1996), the only model so far to take the latter approach is Henson's (in press, a.) SEM model, which does not use a connectionist implementation. Houghton (1994a) suggests improving the positional discrimination of a connectionist CQ model by using an alternative input rule based on a radial basis function (RBF). Radial basis functions (Moody and Darken, 1988, Ballard, 1986;

Kruschke, 1992) can be viewed as generating an elliptical 'receptive field' around a point in multi-dimensional space. If used as the net input function for a connectionist node, such a function can generate a maximal influence from a connection if the input to the connection is numerically equal to the weight on the connection. Higher or lower inputs produce a lower influence. Similar functions are widely used, for example, in competitive learning networks.

While increasing either the dimensionality of the sequencing signal or the power of the activation function would improve the robustness of recall in the model, each approach has its problems. The former involves an increase in the complexity of the sequencing system due to the need for a considerably more complex representation for sequence position. The latter has the problem that, since the improved discrimination is being applied to a very simple timing system, repeated items are difficult to represent. RBFs, for example, have a receptive field with a single centre (or peak response), and repeats would require two or more peaks. An alternative is to make a separation between the occurrence of an item and its identity, such that two different tokens representing two separate occurrences may be bound to the same letter. Clearly this too will lead to an increase in complexity (although other models have taken such a 'type-token' approach, notably those of Page and Norris, submitted, and Dell, 1986, 1988). Additionally, recent evidence from the field of verbal STM suggests that a type-token distinction exists in the processes underlying STM tasks similar to that made here between place-holder (sequence) nodes and content (item) nodes (Henson, submitted). As the RBF idea has yet to be tried out in a full connectionist model this approach is adopted here.

6.1.2 CV Template

The external mode of serial constraint implies the existence of an abstract template, which in the case of a CV constraint will indicate whether the letter in each position of the target word should be a consonant or a vowel. Consonant or vowel letters will then be biased as a group according to the current status of the template. It is not entirely clear what the source of this information might be. Caramazza and colleagues (Caramazza and Miceli, 1990, Link and Caramazza, 1994), in their MOS hypothesis, propose that the explicit representation of CV status forms part of the orthographic

representation used in the graphemic buffer (in the form of CV tags bound to letter tokens). An alternative possibility is that it may be derived on-line from phonology. However, whatever its source, the adoption of an external mode of constraint implies that an abstract CV template aligned with the target word will provide a first-order approximation to the information available to the sequencing system. Accordingly, the present model will be confined to exploring the effect on a CQ spelling system of such an abstract CV template.

6.1.3 Parameters and model complexity

One problem with the CQS model is the large number of parameters governing its operation. An aim for this chapter is thus to simplify the formal basis of the model as far as possible. The major area where such improvements may be made is the interactions between nodes within layers. The equations in the CQS model were largely derived from Houghton's (1990) model, which used lateral inhibition throughout. The need for this outside the competitive filter seems negligible. It is therefore possible to do away with mutual inhibition and self excitation (and their associated parameters) in the bulk of the model. Houghton (1994b, equation 9) has shown that the competitive filter can be made to operate essentially perfectly. The model is therefore further simplified by simulating the operation of the filter. This allows parameters associated with the filter to be removed also.

6.2 Architecture and operation

Figure 6.1 shows the architecture of the new model, CQX[5]. The model comprises four layers of nodes: The sequence, item, letter and filter layers. The item layer nodes represent response *tokens*, while the letter nodes represent response *types*. Letter nodes receive input from all item nodes which are 'bound' to the corresponding letter. The competitive filter receives two new inputs - a CV template biasing competition to be

---

[5] For CQ with external constraints.

between only vowels or only consonants, and a source of noise, added to the filter

rather than directly to letter node activations.

Shifting the locus of noise disruption to the filter removes the obstacles to stopping sequence production using an activation threshold discussed in section 4.2.5. The approach of 4.2.5(b) may now be taken.
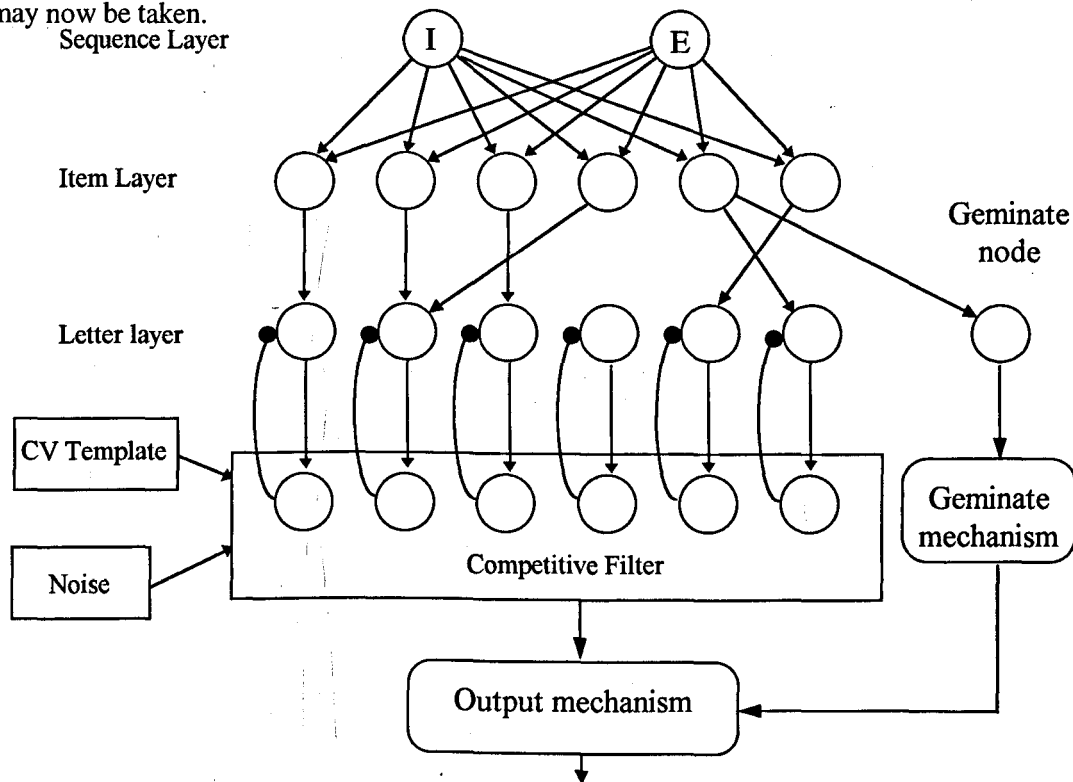
Figure 6.1. *General overview of the CQX model. See text for details.*

Starting at the top of the figure, the sequence layer consists of pairs of I and E nodes

which operate identically to the CQS model (equations 4.2 and 4.3). The item layer

nodes are activated by the I and E nodes using an RBF input rule. Each item node acts

as a place-holder for a different response. During recall, an item node achieves its

maximum activation when the I and E node activations are in the state they were in

when the item was presented during learning. The further the I and E node activations

are from this state, the lower the item node's activation.

The next layer contains letter nodes, each representing one letter. These are activated

by the sequence nodes via excitatory connections with fixed weights, and it is the

arrangement of these connections which defines which sequence of letters will be

produced by the model. Letters may be repeated within a word by connecting two or more item nodes to the same letter node[6].

The most active letter node is selected by a competitive filter, simulated in this implementation by a peak-picking algorithm. As before, the post-production inhibition of response types means that it is not possible to produce a letter twice in a row. Double letters are dealt with by a geminate mechanism similar in structure to that of the previous model, thus there is a geminate node which, when its activation exceeds a threshold, causes the letter currently being produced to be doubled.

During recall a 'CV template' biases the activations in the competitive queue so as to favour consonants or vowels at different points in the recall process. A constant amount of activation is added to all vowel letter nodes if the CV template indicates that a vowel should be produced, or to all consonant nodes if a consonant should be produced. The CV status of each letter in a word is encoded by the model during learning.

6.2.1 Formal description

This section gives a detailed formal description of the model. The simplest possible equations will be used for the model's operation, in order to start from a much simplified formal basis compared with the CQS model.

Item Node Activation

Item nodes are activated by the interaction of the I-E node activation vector with the learned weights in the sequence node to item node pathway according to an RBF activation rule as discussed above. The RBF implemented in the model is an exponential function of the Euclidean distance between the I-E vector and the weight vector to any unit. The net input net_i^I to an item node i from the I-E pair is:

$$net_i^I = e^{-d|s-w|}$$

Equation 6.1

---

[6] This is similar to an arrangement suggested by Milner (1961) in a model of short-term memory.

where **s** is the I-E node vector, **w** is the input weight vector to item node i, and c is a parameter determining the 'slope' of the exponential curve. This parameter governs how sensitive a unit is to the difference between its input weights and the input activation. The higher it is the more finely-tuned the unit becomes, generating a large response only when the input activation matches the weights quite closely. According to equation 6.1, a null activation to a zero weight will generate an input of 1. This is clearly undesirable, so the implemented rule is made conditional on the input weights being non-zero. (Houghton, 1994a, proposes multiplying the radial basis function with an 'energy' component such as a dot-product function to achieve the same effect. This approach however leads to appreciably lower activation levels for items in medial positions, which makes recall less robust than for the approach used here).

|**s-w**| is the Euclidean distance between the vectors **s** and **w**, given by:

$$|\mathbf{s} - \mathbf{w}| = \sqrt{\sum_i (s_i - w_i)^2} \qquad \text{Equation 6.2}$$

Equation 6.1 returns a value with a maximum of 1 when the input and weight vectors are identical and asymptotically approaches 0 as they get further apart.

The activation A_i^I of item unit i is always equal to the net input net_i^I to that item unit; thus the item units are immediately reactive:

$$A_i^I = net_i^I \qquad \text{Equation 6.3}$$

Letter node activation

Binding between item nodes and letter nodes takes place during learning (see below) resulting in one-to-one connections with weights of +1 or 0.

The net input net_i^L to a letter node i from the item units during recall is given by

$$net_i^L = \sum_j A_j^I w_{ji} \qquad \text{Equation 6.4}$$

where A_j^I is the activation of item node j, and w_{ji} is the weight from item node j to letter node i. If a letter is used more than once in a learned sequence it will receive input from more than one item node.

The simplest possible activation function would equate a letter node's activation with its net input. However, a little more is needed to support the CQ approach. As a minimum, inhibited letter nodes need to be removed from the competition for a short time, and this is conventionally achieved as it is in the CQS model by using negative activation values to represent inhibition, and adding a 'decay' term to the activation function in the case of negative activation, thus preventing nodes from recovering instantaneously from inhibition. This results in a function for the activation $A_i^L(t)$ of letter node i at time t as follows:

$$A_i^L(t) = \begin{cases} net_i^L & \text{if } A_i^L(t-1) \geq 0 \\ net_i^L + rA_i^L(t-1) & \text{otherwise} \end{cases} \qquad \text{Equation 6.5}$$

Where r is a parameter which governs the rate of recovery from inhibition.

Notice that in combination with equations 6.1 and 6.4, this will result in a symmetrical activation curve for timesteps before and after the target position for a letter. Thus if a letter is erroneously produced one position too early or too late the activation it will receive by equation 6.5 will be the same in either case. This has implications for the relative likelihood of different types of error. Specifically, it affects the ratio of exchange to deletion errors.

In order to understand this it is necessary to examine factors affecting the deletion – exchange error ratio. Consider a target sequence A B C D E F, and suppose that following the production of A, the letter B which should be produced next has lower activation than C due to noise. C wins, and is output and inhibited. In the next letter position there are two main possibilities: Both B and D will be equally activated according to equation 6.5. If B wins here the result is an exchange error. If D wins the result is, locally, a deletion, and if in all subsequent positions the deletion 'ripples through' the result is a sequence with one deleted letter. Which of these two

possibilities happens is governed by the relative activation levels of the letters preceding and following the current position. If equation 6.5 is modified such that the preceding letter is more active, exchange errors will be more likely. If a modification is such that the following letter is more active, deletion errors will predominate.

It is universally observed that deletion errors form a large part of the error corpora of GBD patients, often with a somewhat higher rate then exchange errors. Deletion errors also predominate in normal speller's 'slips of the pen'. This indicates that an asymmetric activation function may be required for the letter nodes, and initial trials with the symmetrical letter node activation function given by equation 6.5 resulted in the production of a large proportion of exchange errors but no deletions at all. A further term was thus added to equation 6.5 with the purpose of biasing the activation function somewhat towards later rather than earlier letters. For this purpose the gain control term used in equation 4.4 of the CQS model is suitable. Its effect is to reduce the influence of top-down excitation when the target node has been highly active previously. When two letter nodes are simultaneously active this causes the one which has been active for longer (i.e. the earlier letter) to be inhibited relative to the later letter, which has been active for less time.

A similar term is thus added to equation 6.5 as follows:

$$A_i^L(t) = \begin{cases} \left(1 - gA_i^L(t-1)\right)net_i^L & \text{if } A_i^L(t-1) \geq 0 \\ \left(1 + gA_i^L(t-1)\right)net_i^L + rA_i^L(t-1) & \text{otherwise} \end{cases}$$

Equation 6.6

Where g is a parameter $(0 < g < 1)$ which controls the degree of gain control, and thus asymmetry, in the function.

A drawback to introducing an asymmetry biased towards the end of the word is that this is the opposite direction of bias to that which favours insertion errors, by analogous logic to that outlined for deletions. In fact, as well as producing no deletions, the model also produced a relatively low level of insertions (below 10%) when using equation 6.5. However, biasing the activation function towards the beginning of the word (in the manner described below for the geminate node) had only a small effect on the rate of

insertions. Clearly the function can only be biased in one direction, and the bias towards the end of the word is the most beneficial manipulation. Following this manipulation insertion errors are still made, although at a lower rate.

Competitive Filter

The competitive filter nodes receive two additional inputs as well as the main one-to-one excitatory input from letter nodes. First the activation levels of either consonants or vowels are all increased by a fixed amount depending on the current state of the CV template. Secondly, random noise is added to make the process of selecting the most active node prone to error. Initial activation of a filter node A_i^F is thus given by:

$$A_i^F = A_i^L + CV_i + v \qquad \text{Equation 6.7}$$

where CV_i is the CV template input to the node, and v is the noise. In the current implementation the filter is simulated. The most active filter node is assumed to be the winner, and the corresponding letter node is subsequently set to a standard negative (inhibited) value (determined by the parameter _Inh_).

Representation of Geminates

The same basic formulae are used for net input and activation of the geminate node as for the letter nodes. The net input to the geminate node net^G is given by:

$$net^G = \sum_i A_i^L W_{iG} \qquad \text{Equation 6.8}$$

Where W_{iG} is the weight from letter node I to the geminate node (these weights are all set appropriately to either 1 or 0 during learning).

The simple form of activation function shown in equation 6.5 is used for the geminate node. The activation $A^G(t)$ of the geminate node at time t is thus:

$$A^{G}(t) = \begin{cases} net^{G} + v & \text{if } A^{G}(t-1) \geq 0 \\ \\ net^{G} + rA^{G}(t-1) + v & \text{otherwise} \end{cases}$$ Equation 6.9

Where r is the same recovery rate parameter as that used for letter nodes, and v is the same random noise value used in equation 6.7. Following successful triggering the geminate node is inhibited in the same way as letter nodes.

The mechanism is capable of producing multiple geminates in words such as 'bookseller', but this requires fairly sensitive setting of the threshold for geminate node firing and the recovery rate from geminate node suppression. The mechanism is incapable of producing double geminates, which do occur very occasionally as in the word *balloon*, for the same reason that the basic CQ model is unable to produce double letters unaided - the inhibition of the geminate node after firing, which is necessary to prevent perseveration. However, chunking or grouping effects may be important in words with multiple geminates. To avoid biasing the results due to these limitations, words containing more than one geminate were excluded from the test corpus for the simulations reported here.

CV Template

The model requires that information regarding the C/V status of each target letter be reliably available. In order to demonstrate a concrete example of a working template the following arrangement is implemented in the model. However no theoretical weight is claimed for this particular implementation.

An additional pair of nodes, *Cnode* and *Vnode*, are excited by the item nodes via weights which are set to 0 or 1 during the learning process. For each position, if the corresponding letter is a vowel the weight from that item node to the Vnode is set to 1, and that to the Cnode is set to 0, and *vice-versa* for consonants. This may be achieved with a simple Hebbian procedure discussed below. During recall, the template nodes are activated according to the following equations:

$$Cnode = \sum_i \left(A_i^I W_i^C\right)^4 \qquad\qquad \text{Equation 6.10}$$

$$Vnode = \sum_i \left(A_i^I W_i^V\right)^4 \qquad\qquad \text{Equation 6.11}$$

Where W_i^C and W_i^V are the weights from item node i to the C and V nodes respectively. The state of the template is then determined by taking the most active of the two nodes. The activation-weight product is raised to the fourth power in order to provide a sufficiently sharply peaked signal that sequences such as "CCCVCCC" can be accurately distinguished.

Learning and Production

Sequences are learned by first setting the sequence-to-item weights, and then binding each item node to a letter node. The target sequence is presented to the model letter by letter, each input simultaneously activating, to a value of 1.0, a letter node, an item node representing the individual "event" or token, and optionally the geminate node. During presentation of the letter sequence, the I- and E-nodes obey equations 4.2 and 4.3 respectively. As each item is presented Hebbian learning takes place in the sequence-to-item connections - that is, the weights are set to the product of the item activation and the current I- and E-node activations. Since item activations are either 0 or 1, this reduces to copying the state of the I-E vector into the weight vector to the active item node. The binding of the item node to the current letter identity and geminate node (and of C and V template nodes to item nodes in the suggested implementation of the CV template) is achieved using the same rule. All learning in the model is thus unsupervised.

During spelling production, the I and E nodes again obey equations 4.2 and 4.3. Item nodes are activated according to equation 6.1, and activate their associated letter nodes by equation 6.6. The input to the competitive filter is given by equation 6.7. The competitive filter generates an output by selecting the most active of the competing responses, and the corresponding letter node is then inhibited. A 'stopping' threshold T_S is applied to the activations of the letter nodes. When no letter node activation exceeds T_S, the spelling is deemed to be complete and the spelling operation stops.

6.2.2 Fitting the model to patient performance

The aim for this model is not to perfectly match the performance of a particular GBD patient - indeed it would be surprising if a model of the serial output process, disrupted only by the addition of unstructured noise, were able to closely match performance in an area such a spelling, which is surely influenced by a variety of idiosyncratic and strategic factors over and above basic sequential mechanisms. The aim in these simulations is therefore to produce a "basic generic GBD" performance. In contrast with the CQS model the CV structure of the stimulus words affects performance, so real words are used in the test corpus as detailed below.

Table 6.1 summarises the default parameter values which were used in all simulations unless otherwise stated.

Parameter	Symbol	Value
Decay rate for I- and E-nodes	δ	0.8
Slope of item activation function	c	4.0
Letter node gain control	g	0.6
Letter node recovery rate	r	0.7
Letter node inhibition level	Inh	-1.0
Geminate node trigger threshold	T_G	0.8
Stop threshold	T_S	0.2
C/V template activation bias	B_{CV}	0.2
Noise magnitude	v	±0.422

Table 6.1. *Default parameter values used in all simulations unless otherwise stated.*

The model contains nine parameters. This is a low number compared with the CQS model, and moreover several parameters are quite specific in their effects. It is thus entirely possible to set the parameters by hand to match the model's performance to data, which is the approach taken here. As this approach is open to the criticism of

being to some extent ad-hoc, it is most important to address the issue of the parameter-dependence of the model - in other words, to what extent the results have been prejudiced by explicitly setting the parameters to give the 'correct' performance. This is addressed through an extensive parametric test of the model, reported in study 9 of the next section.

6.3 Simulations

6.3.1 Corpus

The simulations in this chapter were all carried out on a corpus of English words taken from the Oxford Psycholinguistic Database (Quinlan, 1993). As many of the words in the database are long (and infrequent) the model was not expected to be able to spell them all perfectly in the absence of noise. Longer sequences are more difficult for a CQ system to produce, and repeated items are particularly fragile, making production of long polysyllabic words difficult. However chunking mechanisms may play some part in the production of polysyllabic words (Hartley and Houghton, 1996). The model was tested on all words in the database with one or fewer double letters, in the absence of noise. About 1% of the words could not be spelled correctly, all of which were of 9 letters or longer, with the exception of a small number of eight-letter words.

4437 words of six letters with at most one double letter formed the core of the test corpus, as much of the detailed error analysis is again carried out on six-letter words for easy comparison with experimental data. To these were added a randomly chosen subset of words of different lengths, excluding those with more than one double letter and those which the model could not spell in the absence of noise. Table 6.2 shows the composition of the test corpus.

The data reported below were gathered from several simulations. In each case data was collected during a number of passes over the full test corpus of 8294 words. Most simulations in this section used forty passes over the test corpus. The parametric investigation reported in study 9 used two passes at each of a large number of parameter combinations.

Word length (letters)	Number of words	Number of geminate words
1	3	0
2	11	0
3	95	2
4	292	30
5	444	64
6	4437	858
7	680	140
8	666	124
9	676	126
10	485	75
11	311	52
12	194	20

Table 6.2. *Composition of the test corpus.*

Study 1 - Basic Performance

This first study examines the basic performance of the model under the influence of noise disruption. Figure 6.2 shows the overall percentage of correct spellings produced by the model for each word length in the test corpus at each of five values for the noise magnitude. As expected, and as for the CQS model, longer words and higher noise both render spelling more error prone. The noise values required to produce a similar level of disruption are an order of magnitude higher than those for CQS, demonstrating that the higher positional discrimination of the RBF input rule makes spelling considerably more robust.
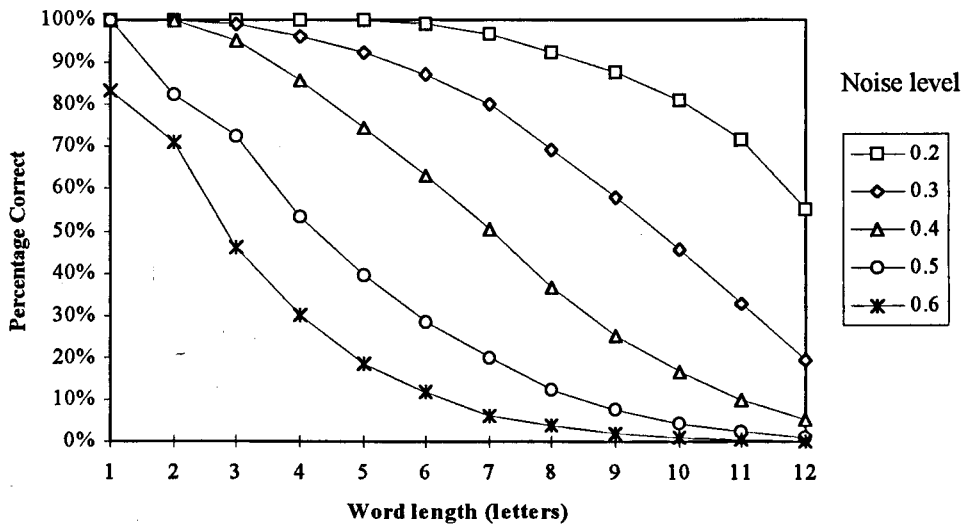
Figure 6.2. *Performance of the CQX model plotted against word length for five different levels of random noise.*

Figure 6.3 compares the performance of the model with that of two GBD patients, AS (Jonsdottir et al., 1996) and JH (Kay and Hanley, 1994), using two different noise settings. Some care was taken to find the noise values which give the best fit, to enable an accurate comparison to be made between the model and the data. It is important to point out that the fit obtained in Figure 6.3 was not achieved by manipulating any parameters other than the noise level. The other parameters of the model were adjusted only to obtain a qualitative match to other aspects of patient performance on 6-letter words, where the majority of data on more detailed aspects of error performance is available.

To make the simulations more concrete, performance was matched to that of a specific patient, AS. This patient is English, which is important as the test corpus uses English words, and is one for whom a good deal of relevant data is available. A noise level of 0.422 provides a close fit to AS's performance on 6-letter words, so this value was chosen as the standard for all simulations. As Figure 6.3 shows, a good fit was also obtained to patient JH with a lower noise value of 0.32.
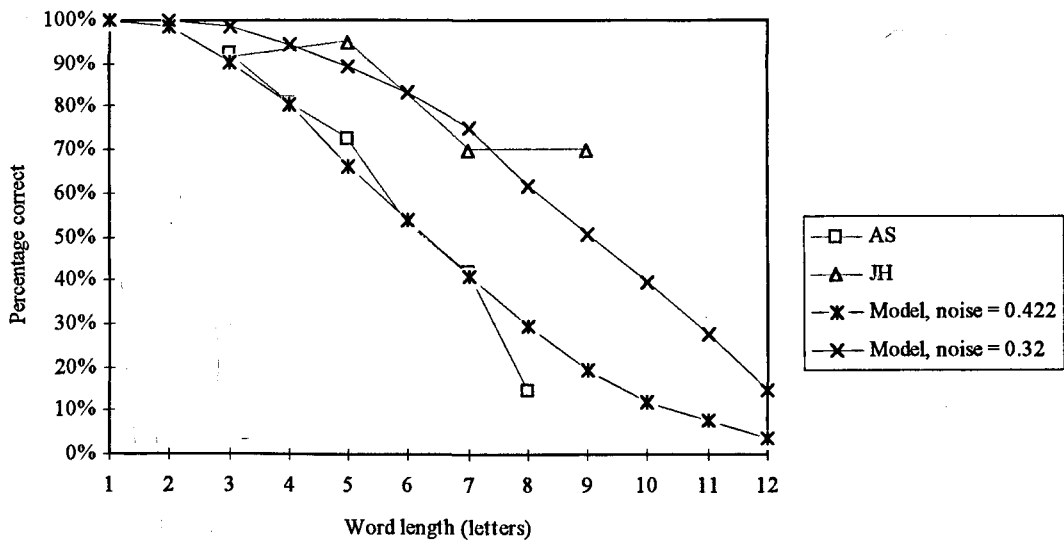
Figure 6.3. *The performance of the model compared with that of GBD patients JH and AS.*

The mismatch with the data point for 8-letter words is interesting, as a similar trend is shown by both AS and LB, indicating that this is not merely due to a low sample size, for example. Strategic effects such as chunking are presumably more likely at longer word lengths, so it is difficult to tell if this mismatch is important.

Comparison with CQS model

With the exception of the point for 8 letter words, the 0.422 curve of Figure 6.3 matches the performance of AS quite closely. Figure 6.4 compares this fit to that obtained with the CQS model (in Figure 4.4). The fit obtained by the current model is clearly an improvement.
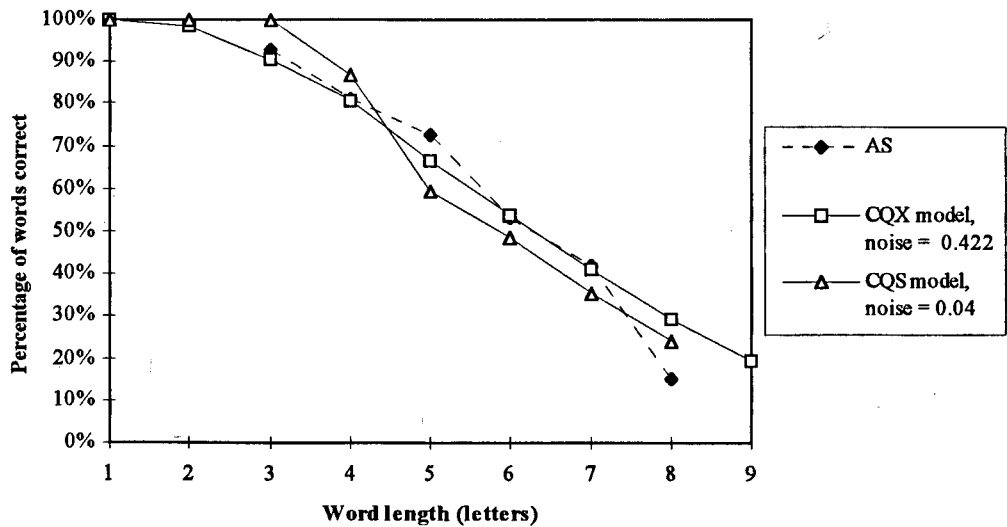
Figure 6.4. *The fit between the model's performance with a noise level of 0.422 and that of AS, compared with that given by the CQS model with a noise level of 0.04.*

A major factor in the difference between the two models is that the CQS model requires a phase of supervised practice on longer words. The model requires no practice at all for words up to three letters in length, however, and only very minimal practice for four letter words. These shorter words are thus learned under a qualitatively different regime to longer words, and as can be seen from Figure 6.4 they are more robust in the face of noise than would be expected by the performance on longer words, leading to a kink in the curve between words of four and five letters. The current model is able to apply the same single-step learning procedure to words of all length as a result of the greater robustness of the RBF input rule approach, and this results in a smoother relationship between performance and word length.

Study 2: Error incidence.

The scoring procedure and automatic error analysis algorithm discussed in Chapter 4 were again used to analyse the errors made by the model on six-letter words. Figure 6.5 shows the overall serial position curve for single-error responses. Again, the model shows a higher incidence of errors in medial positions than at the start and end of the words. The reason for this is the same as with the CQS model: the I-E timing signal changes more rapidly near the start and end of the sequence than in medial positions,

and thus in the middle of the word consecutive letter positions are less well discriminated, even using the RBF-based input rule, than those at the start or end.
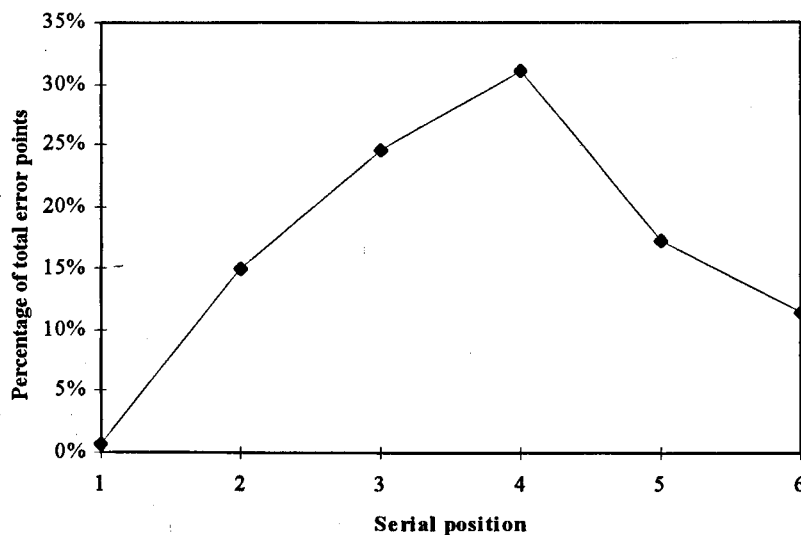
Figure 6.5. *The incidence of single-error responses made by the model at each serial position in six-letter words, given in terms of the error points awarded by the scoring procedure of Caramazza and Miceli.*

As with the CQS model the curve is biased towards the end of the word compared with the patient data. However, Figure 6.6, which compares the serial error curves produced by both the CQS and current models with that of AS, shows that the disparity is considerably reduced with the CQX model.
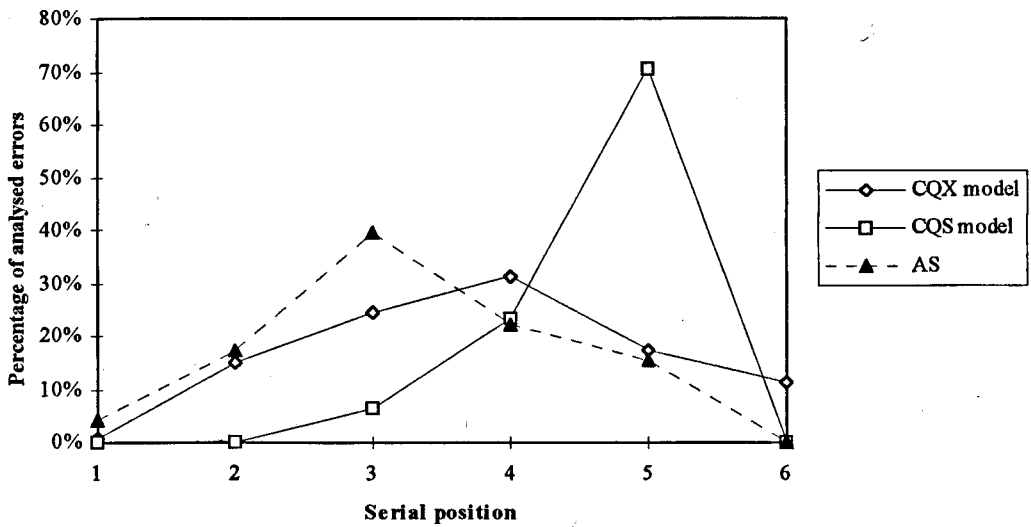
Figure 6.6. *Comparison of serial error incidence curves produced by the CQS and CQX models with that of GBD patient AS.*

Looking in more detail at the nature of the errors, Figure 6.7 shows the serial error curves from the model individually plotted for the five major error types.
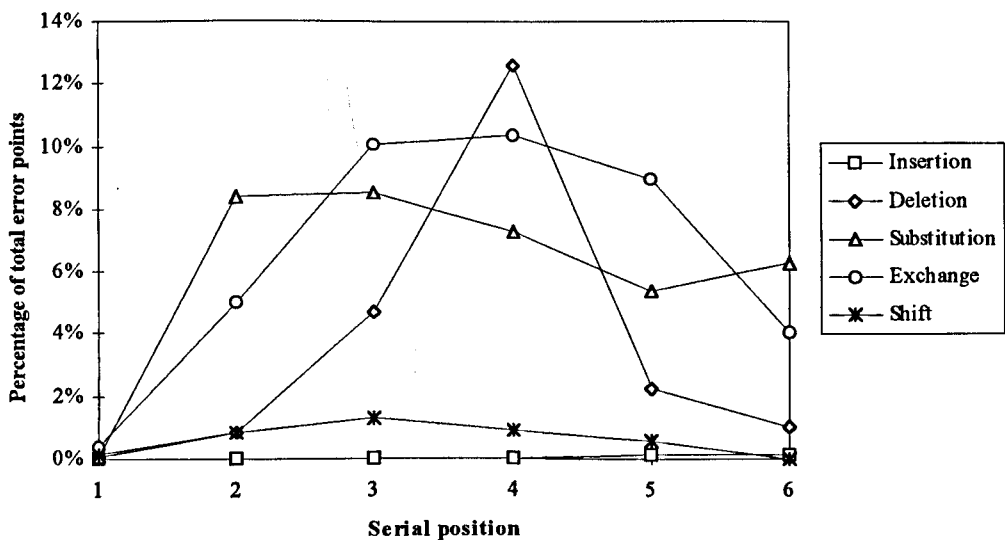
Figure 6.7. *Serial error curves produced by the CQX model on six-letter words, plotted separately for each of the five basic error types.*

The reason for the considerably less skewed error curve compared with that of the CQS model is immediately evident: Deletions, substitutions, shifts and exchanges all peak medially, in contrast to the highly skewed curves for deletions and substitutions in the CQS model (Figure 4.6). The model produces a varied mix of errors, with no evidence of the stereotyped substitution errors of the CQS model. This improvement is due to the RBF activation rule, which gives a smooth increase in letter activation before the target position and a smooth fall-off after it, reducing the tendency in CQS for letters to be re-activated towards the end of words, and leading to more symmetrical serial error incidence curves.

Figure 6.8 shows the comparable data for GBD patients LB (a) and AS (b). Interestingly, substitutions show a 'tail' of high incidence towards the end of the word both for AS and for the model. AS shows insertions increasing to a peak in the final position, but with a subsidiary peak in the middle of the word, and the model also shows this feature of a raised insertion rate towards the end of the word. However, the low rate of insertions in the model compared with that for the patients suggests that this type of error is not fully explained on the model.

(a)

(b)
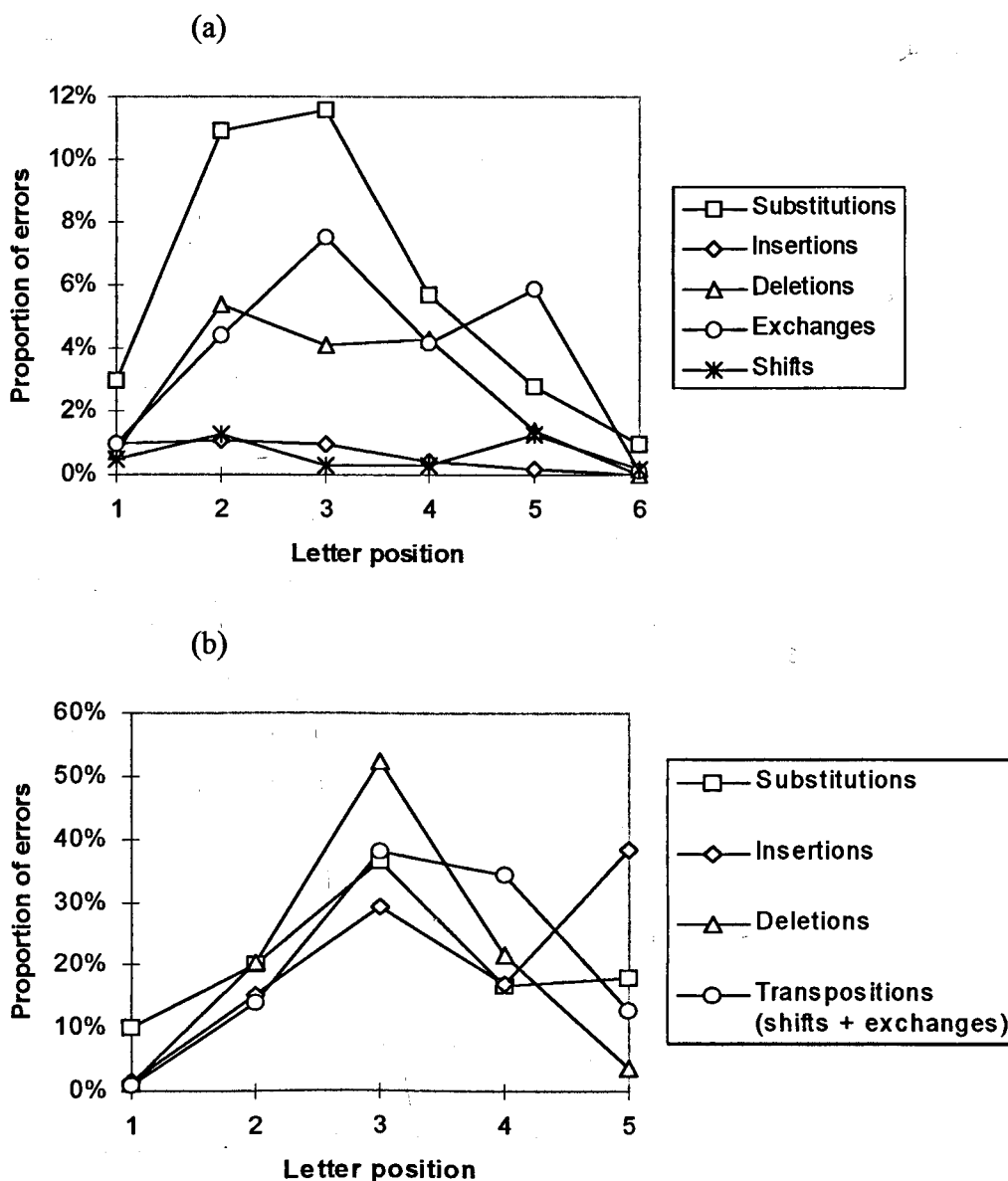
Figure 6.8. *Serial error curves on six-letter words, plotted separately for each of the five basic error types, for GBD patients AS (a) and LB (b).*

Study 3: Error types and proportions

Figure 6.9 compares the overall incidence of the different error types produced by the model with those of GBD patients AS, LB, JH and HE. The same general caveats introduced in Chapter 4 apply to this comparison: The patients themselves are somewhat variable in the pattern of error incidences, and matching the pattern of any particular patient accurately was not an aim of the model. However, given that the only effort made to directly influence this distribution was the introduction of the gain

control term in the letter activation rule, without which the model produces no deletion errors at all, the fit of the model with the general trend in the data is encouraging. This aspect of the model's behaviour will be returned to in study 9, where the effect of varying specific parameter values is examined.
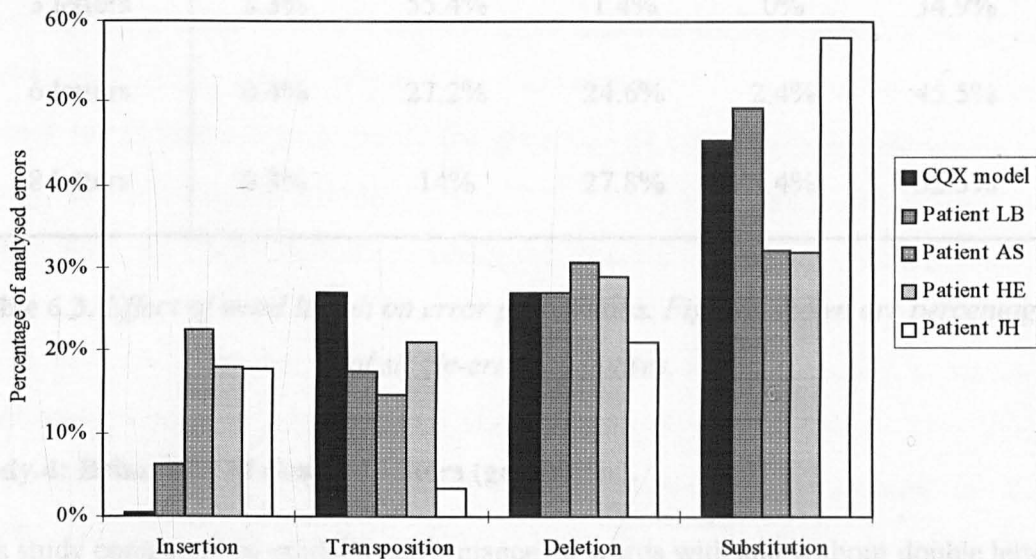
Figure 6.9. *Relative proportions of five major error types for the model on six-letter words, compared with four GBD patients LB, AS, JH and HE. Shift and exchange errors are combined under 'transpositions'.*

Finally, an apparent similarity across a number of GBD patients is that the relative proportion of deletion errors increases with increasing word length, while that of substitutions decreases. Table 6.3 shows the effect of word length on error proportions for the model. The model shows the opposite trend to that apparent in the majority of GBD patients, with deletion errors decreasing and substitution errors increasing as a proportion of all single-error responses with increasing word length. Clearly at this level of detail the error mechanisms in the model do not fit well with the experimental evidence. However, as with Figure 6.9, this data must be treated with care - there is some variability between patients, the proportions shown are for single-error responses only, rather than single-*type* errors, and it is not clear for most patients how the trend in relative incidence relates to the absolute incidence of the various error types.

Nonetheless, this mismatch will be returned to in section 6.4, where an alternative mechanism for deletion errors will be proposed which may address this problem.

Word length	Insertions	Deletions	Exchanges	Shifts	Substitutions
3 letters	8.3%	55.4%	1.4%	0%	34.9%
6 letters	0.4%	27.2%	24.6%	2.4%	45.5%
8 letters	0.3%	14%	27.8%	5.4%	52.5%

Table 6.3. *Effect of word length on error proportions. Figures shown are percentages of single-error responses.*

Study 4: Behaviour of doubled letters (geminates).

This study compares the model's performance on words with and without double letters ('geminate' and 'non-geminate' words). Internally, a geminate word is treated as a word one letter shorter and without the doubling, with separate concurrent representation of the doubling ('geminate feature') by the triggering of the geminate node at the appropriate point in the spelling. For the letter sequence itself the error rate will thus be less than would be expected, being consistent with that of a word one letter shorter than the target word. However, geminate words are subject to a further source of error not present in non-geminate words: The geminate node is itself subject to noise, and does not always trigger correctly. The additional error rate due to noise on the geminate node and the lower error rate intrinsic to the shorter letter sequence counteract each other to some extent, and the overall error rate on geminate words depends on the relative strengths of the two effects.

In all other simulations the noise applied to the geminate node has the same magnitude as that applied to letter nodes. It is interesting to examine the effect of varying the relative noise levels on letter and geminate nodes, however. Figure 6.10 plots the percentage correct geminate words against word length for various values of the parameter 'geminate noise factor', which acts as a multiplier on the level of noise in the

geminate system. With a geminate noise factor of 1.0 the noise applied to the geminate node is equal to that in the competitive filter; lower values correspond to lower values of geminate noise relative to competitive filter noise. The performance of the model on non-geminate words is included for comparison. Figure 6.10 shows (broken lines) that with the same magnitude of noise applied to the geminate node and letter nodes the error rate on geminate words is slightly higher than that for non-geminate words of the same length. GBD patients vary in their performance on this comparison. The English patient AS produces approximately the same rate of errors on geminate words as on non-geminate words (Jonsdottir et al., 1996), while Caramazza and Miceli (1990) found that their Italian patient LB performed considerably better on geminate words than on non-geminate words of the same length, at least for 6-letter words. Jonsdottir et al. provide a number of arguments to suggest that differences in details of performance between the Italian LB and English AS may be due to LB's spelling benefiting to some degree from the greater sound-spelling regularity of Italian compared to English. In this connection it is instructive to note that in 'standard' Italian doubled consonant letters always correspond to a phonological difference, whereas in English this is not so, e.g., *s* in *this* vs. *miss*, *m* in *command* vs. *demand*; *e* in *scene* vs. *screen*. This potential phonological cueing as to the position of a doubled letter could make the geminate node less subject to noise in Italian than in English.

There is thus no a priori reason to suppose that the vulnerability of the geminate node to noise disruption is the same as that of letter nodes. Varying the 'geminate noise factor' parameter allows a more robust geminate system to be simulated, and Figure 6.10 shows (solid lines) that with a value of 0.8 of the standard noise level performance on geminate words closely matches that on non-geminate words, as is the case for AS, while with a lower value of 0.5 (and thus correspondingly more robust geminate mechanism) performance on geminate words approximately matches that for non-geminate words one letter shorter in length, which is the case with LB on 6-letter words (the only length for which there is precise data).
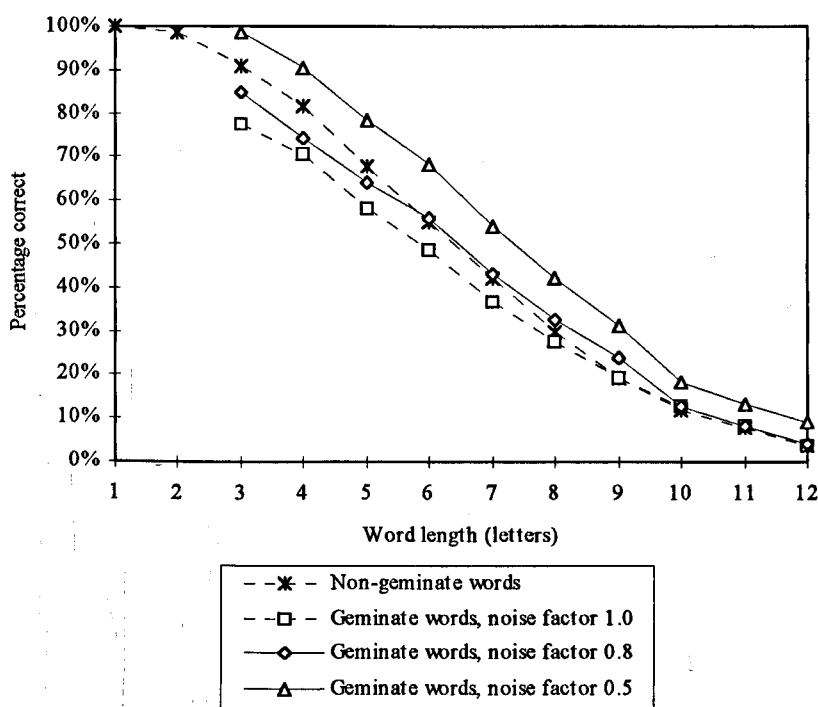
Figure 6.10. *Performance of the model on non-geminate words compared with performance on geminate words with varying degrees of noise in the geminate system. See text for details.*

Chapter 4 reviewed the characteristics of GBD errors which appear to result from errors in the production of a 'geminate feature'. Table 6.4 summarises these characteristics, and compares the errors made by the model in triggering the geminate mechanism.

By contrast with the CQS model, the CQX model produces all the geminate error types seen in GBD patients, the higher positional resolution allowing geminate feature insertion and deletion errors. A quantitative fit to the data in this area was not an aim of the model. However, some qualitative points emerge. The majority of errors involve the movement or deletion of the geminate feature in the model, which is in agreement with the data. The patients make essentially no errors in which a double letter is introduced into a non-geminate word, and the model agrees on this point also.

In the model these effects are straightforward consequences of the geminate mechanism. The inhibition of the geminate node after it has triggered makes it unlikely

that it will be triggered more than once in the same word, hence the rarity of new geminates appearing in words which already contain one. In non-geminate words the geminate node is completely inactive in the absence of noise, and the presence of noise alone is not sufficient to trigger it; hence the lack of geminates appearing in non-geminate words.

Error type	Patient LB (n = 69)	Patient AS (n = 50)	Model (n = 12650)
The 'geminate feature' itself shifts. (e.g. *abisso* → *abbiso*, patient LB)	36%.	30%	23%
The doubling occurs in the correct position, but the wrong letter is doubled. (e.g. *sorella* → *solerra*, patient LB)	27%	Not listed	12%
The 'geminate feature' does not occur. (e.g. *missile* → *misile*, patient AS)	13%	30%	63%
A new geminate feature is introduced into a word which already contains a doubling. (e.g. *pepper* → *peepper*, patient AS)	18%	4%	2%
A new geminate feature is introduced into a word with no doubling. (e.g. *tavolo* → *tavvolo*, patient LB)	0.02%	0%	0%

Table 6.4. *Percentage of errors involving doubled letters falling into five classes.*

Study 5 - Effect of lexicality

The effect of lexical status is better modelled in the CQS model by variation in noise level than by the level of 'over-learning' in the model. The CQX model does not include an incremental learning process, so this manipulation is not directly applicable. Accordingly, lexical status will be modelled here by manipulating the noise level, a higher noise level being taken to model performance on non-word stimuli. Figure 6.11 shows the overall performance of the model with a slightly higher noise level of 0.46 compared with the standard level of 0.422. As expected, performance is uniformly worse for the higher noise level. GBD patients show a qualitatively similar error pattern for word and non-word stimuli, so Figure 6.12 and Table 6.5 compare the relative proportions of the different error classes and the geminate error pattern, respectively, for standard and high noise conditions.
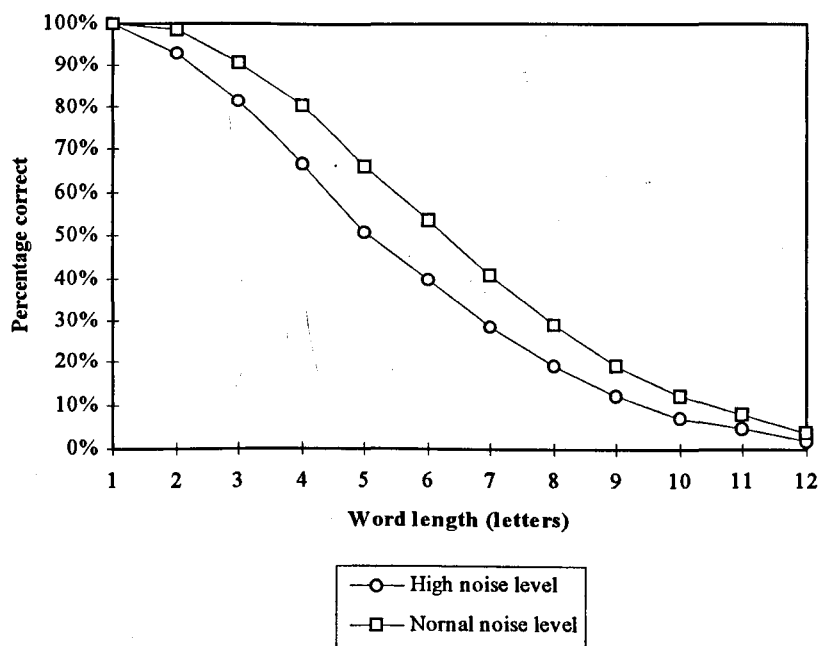
Figure 6.11. *Performance of the model plotted against word length for standard (± 0.422) and high (± 0.460) levels of noise.*
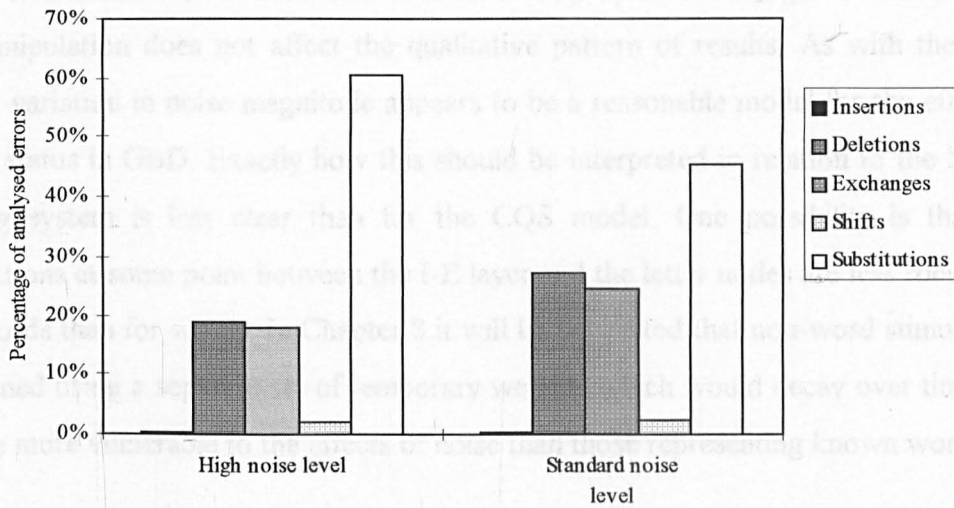
Figure 6.12. *Relative proportions of five major error types made by the CQX model on six-letter words, for standard (± 0.422) and high (± 0.460) levels of noise.*

Error type	Model Standard noise level	Model High noise level
The 'geminate feature' itself shifts.	23%	27%
The doubling occurs in the correct position, but the wrong letter is doubled.	12%	18%
The 'geminate feature' does not occur.	63%	52%
A new geminate feature is introduced into a word which already contains a doubling.	2%	3%
A new geminate feature is introduced into a word with no doubling.	0%	0%

Table 6.5. *Percentage of errors involving doubled letters falling into the five classes of Table 6.4, for standard (± 0.422) and high (± 0.460) levels of noise.*

Figure 6.12 and Table 6.5 show that both for error proportions and geminate behaviour the manipulation does not affect the qualitative pattern of results. As with the CQS model, variation in noise magnitude appears to be a reasonable model for the effect of lexical status in GBD. Exactly how this should be interpreted in relation to the human spelling system is less clear than for the CQS model. One possibility is that the connections at some point between the I-E layer and the letter nodes are less robust for non-words than for words. In Chapter 8 it will be suggested that non-word stimuli may be learned using a separate set of temporary weights which would decay over time and thus be more vulnerable to the effects of noise than those representing known words.

Study 6 - Effect of CV information on spelling dynamics

The previous simulations have so far demonstrated that the basic performance of the CQS model is preserved in the CQX model, and is improved upon in several areas. The next three studies examine the effect of providing CV status information to the spelling process. It is instructive before looking at error data to see how the CV input affects the dynamics of the spelling process. To do this Figures 6.13(a) and 6.13(b) show the time course of letter activations during the spelling of the word 'CINEMA', with and without influence from the CV template. The graphs show the activation levels of each letter in the word at each timestep during spelling production, and the activation levels of consonants and vowels not present in the target word. The CV bias level used in Figure 6.13(b) is somewhat higher than that used in the other simulations in order to show the effect of the parameter clearly.
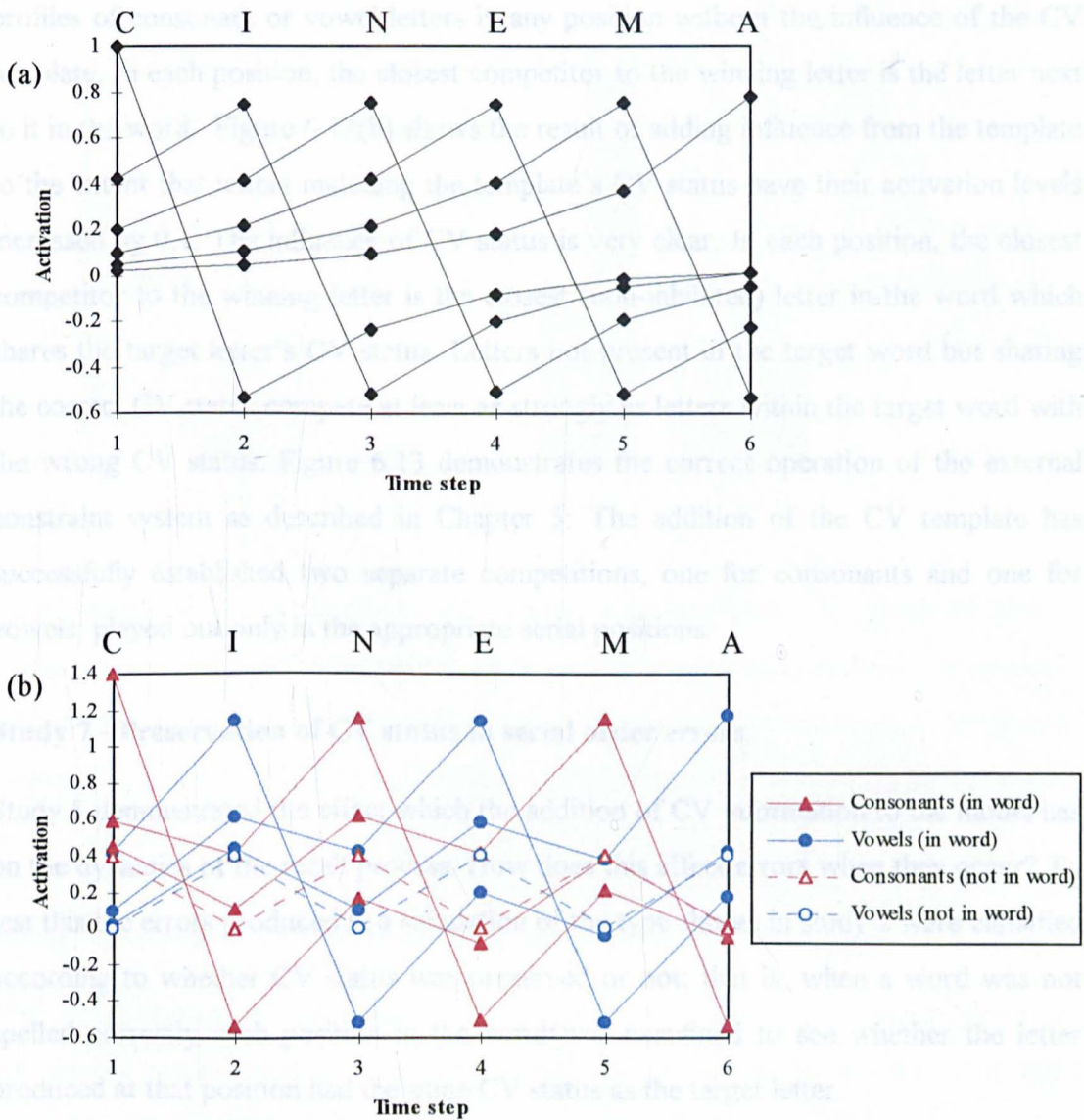
Figure 6.13. *The effect of CV template information on the process of spelling the word 'CINEMA'. (a) shows the activations of each letter in the word at each time step in the absence of CV information. (b) shows the same process with the CV bias set to 0.4. The curve for each letter is labelled at the point where that letter 'wins' the output competition and is output.*

Comparing the two graphs, Figure 6.13(a) shows the dynamics usually seen in CQ models, with the activation of each letter rising to a peak at the point where it wins the output competition, then falling to a negative value as it is inhibited, followed by a slow recovery. As expected, no qualitative difference is evident between the activation

profiles of consonant or vowel letters in any position without the influence of the CV template. In each position, the closest competitor to the winning letter is the letter next to it in the word. Figure 6.13(b) shows the result of adding influence from the template to the extent that letters matching the template's CV status have their activation levels increased by 0.4. The influence of CV status is very clear. In each position, the closest competitor to the winning letter is the closest (non-inhibited) letter in the word which shares the target letter's CV status. Letters not present in the target word but sharing the correct CV status compete at least as strongly as letters within the target word with the wrong CV status. Figure 6.13 demonstrates the correct operation of the external constraint system as described in Chapter 5: The addition of the CV template has successfully established two separate competitions, one for consonants and one for vowels, played out only in the appropriate serial positions.

Study 7 - Preservation of CV status in serial order errors.

Study 5 demonstrated the effect which the addition of CV information to the model has on the dynamics of the recall process. How does this affect errors when they occur? To test this the errors produced in a simulation of the type shown in study 2 were classified according to whether CV status was preserved or not; that is, when a word was not spelled correctly each position in the word was examined to see whether the letter produced at that position had the same CV status as the target letter.

The results are shown in Figure 6.14, which compares single-error responses in which CV status was preserved with those which violated CV constraints for the key error types of transposition (exchange and shift errors combined) and substitutions. In both cases the CV template is clearly constraining errors, with the majority of errors preserving CV status. Overall 80.3% of transpositions and 97.2% of substitutions preserved their CV status, in both cases rather higher than AS's rate of 70.8% and 85% respectively, but showing the same trend of better CV preservation for substitutions than transpositions.
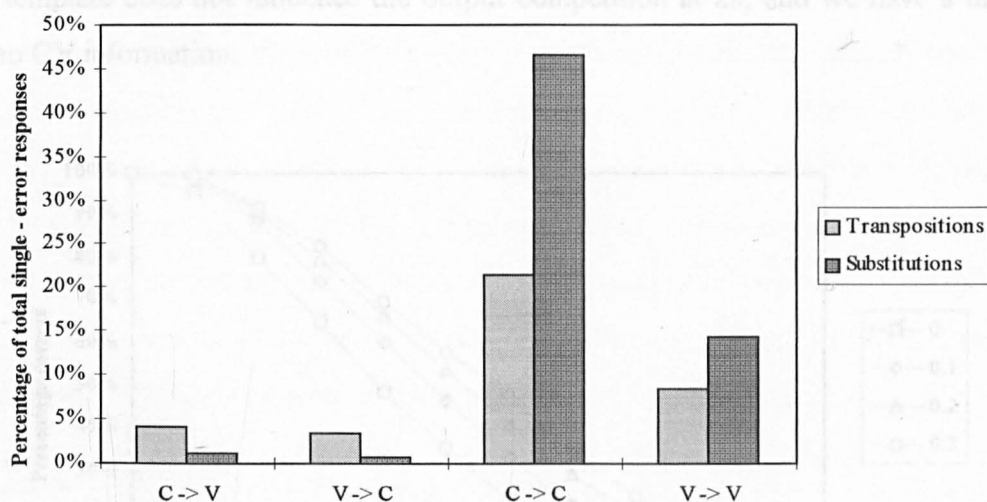
Figure 6.14. *The preservation of CV status in spelling errors by the model. The proportion of errors is shown which involve the changing of a consonant to a vowel (C → V), a vowel to a consonant (V → C), a consonant to a consonant (C → C) or a vowel to a vowel (V → V), in both substitution and transposition (shift and exchange) errors.*

Study 8 - Utility of CV information during spelling production.

The incorporation of consonant-vowel status information in the model allows the preferential preservation of consonant/vowel status observed in the errors of GBD patients to be modelled. The incorporation of such information is driven by empirical evidence, however, and the question arises whether the availability of this information to a spelling system has any practical utility - in other words, is there any evidence that spelling performance is improved by such information? Chapter 5 has suggested that an external constraint system should act to reduce errors, and if this were the case it would provide a functional rationale for why such information appears to be used.

The strength of the input from the CV template to the letter selection mechanism is a parameter of the model, and hence can be varied. Figure 6.15 shows the effect of the CV bias parameter on the performance of the model, again plotted as word length vs. percentage correct. This time all the curves are plotted for the standard noise magnitude of 0.422, but with CV template bias levels between 0 and 0.3. With the bias level set to

0 the template does not influence the output competition at all, and we have a model with no CV information.
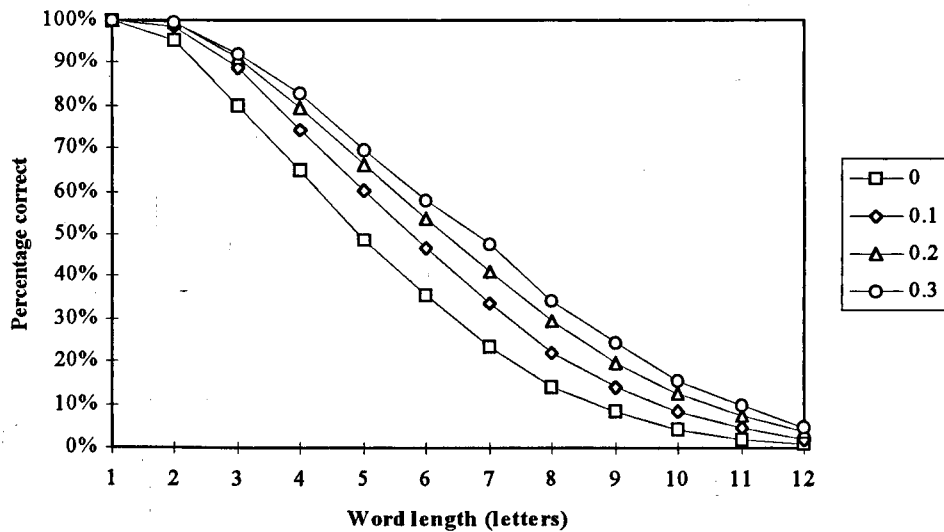
Figure 6.15. *The effect of varying the CV bias parameter on the performance of the model on words of different lengths.*

Clearly, increasing the level of template influence from 0 to 0.3 improves the accuracy of spelling in the face of the same level of disruption from background noise. Intuitively, the effect of the bias is to weaken the degree of competition from letters of opposite CV status, while maintaining the same level of competition from letters of the same CV status. The overall effect of this depends on the CV structure of individual words, and changes from letter to letter. Figure 6.15 demonstrates that the availability of CV information during spelling production is functional for this model, and hence that the model provides a theoretical basis for understanding why it might be used.

Study 9 - Parameter dependence of model behaviour

The CQX model contains eight free parameters (plus the noise level) which have been set manually. As with the CQS model, the parameter settings have in general been fixed for all studies. However, as discussed in Chapter 4, it is most important to check the sensitivity of the model to parameter variation. With its lower number of parameters than CQS, it is possible to carry out a comprehensive study of a large portion of the

model's parameter space. Six parameters were chosen for this study, of which 2 affected the inherent accuracy of the positional cueing of letter activation by the start-end context signal (parameters δ and *c* in Table 6.1); another 3 affected the detailed dynamics of the letter node activations (parameters, *g, r* and *inh* in Table 6.1); and the sixth the strength of the CV bias (B_{CV}). Each parameter was varied in 5 steps between half and twice its default value. With six parameters this produces $5^6 = 15,625$ parameter sets, each tested on two complete passes through the test set of 8294 words, a total of nearly 260 million runs of the model.

Parameters *c, inh* and B_{CV} are varied by multiplying their values by 0.5. 0.75, 1.0, 1.5 and 2.0. Parameters δ, *g* and *r* may only vary between 0 and 1, since they specify decay rates. These parameters are thus more reasonably treated by multiplying the difference between the parameter value and 1.0 (since all are greater than 0.5) by the same set of values. Error distributions are best compared when overall performance is as equal as possible, hence for each set of parameter values, noise was normalised to produce as near as possible an overall performance of 56% correct on six-letter words, the level achieved with the default parameter values. In about 4% of cases it proved impossible to achieve this level of performance even without noise, however this was due to extreme parameter values. Over a narrower range of variance, 0.75 to 1.5 times the default values, no cases failed to achieve the baseline performance. The few parameter sets which could not support this level of performance were discarded.

The model's performance was analysed for five major features:

1. Word length effects. An effect was taken to be shown where performance declined monotonically with increasing word length.

2. Serial position effects. The effect required was a lower error rate in initial or final than in medial positions.

3. Ranking of incidence of error types. The factor of interest here is the stability of the gross ranking of error proportions seen with the default parameter values (study 3), with substitutions the most frequent errors, exchanges and deletions intermediate, insertions lower than either and shifts least.

4. Preservation of CV structure in errors. A reliable effect was taken to be shown if the number of errors preserving CV structure was at least 10% greater than those not preserving it.

5. The ranking of geminate error types as compered to GBD patients AS and LB (the only patients for whom this has been studied). The pattern required was that shifts and deletions of the geminate feature should both be more common than the introduction of new geminate features.

Table 6.6 shows the model's performance on each of these criteria over the full parametric study. In addition, it separately reports results over the smaller parameter space between 0.75 and 1.5 times the default value. This provides some indication of the effect of extreme parameter values.

Effect	Percentage of successful runs showing effect	
	Full survey: 0.5-2.0 × parameter default values	Narrow survey: 0.75 - 1.5 × parameter default values
Word length effect	75%	77%
Overall incidence of errors peaks in medial positions	81%	93%
Correct ranking of error type proportions	68%	94%
C/V-status preserved	100%	100%
Correct ranking of geminate error types	63%	69%
All effects except error type ranking and geminate error ranking	63%	72%
All effects	27%	52%

Table 6.6. *Results of parametetric study*

As can be seen, the properties investigated are generally robust in the face of parameter variation, all being individually present in over 60% of cases. The value for the word length effect is artificially low because of statistical variation - only two runs of the model were made with each parameter set, and with the low number of words at the extreme word lengths, combined with the less rapid change in performance at either end of the word length curve (see Figure 6.2) random fluctuations can cause a violation of the strict requirement that performance should fall with each increase in word length. When only 5, 6 and 7 letter words are taken into account, the figure rises to close to 100%.

Properties 1, 2 and 4 of the above list are thus robust, being found in 80% or more of cases. With respect to property 3, ranking of error types, this turns out, to be a little more parameter dependent, a result which accords with more variable performance of the GBD subjects on this measure. Additionally, study 3 has raised some doubts about the appropriateness of the mechanism for deletion errors in the model. However, the ranking is highly stable compared with the CQS model and 94% of cases in the narrow survey still show the default pattern, surprisingly high for such a complex, emergent effect. The ranking of geminate error types is also a less robust result, being correct for 63% of the parameter sets in the full survey.

Table 6.6 shows that good performance is preserved when properties 1, 2 and 4 are taken together - 63% of runs on the full survey and 72% on the limited survey display all three. The total falls considerably when the number of runs on which all five properties are shown is assessed. This is due to many of the parameter sets which display property 3 failing to display property 5, and vice-versa. Nevertheless, the entire error pattern is shown by over half the cases in the narrower survey.

In conclusion, most major features of GBD are robust properties of the model under noise, and the more variable features in GBD patients' performance, ranking of error types and behaviour of geminated words, show an interesting degree of parameter dependence in the model.

6.4 Discussion

There were two main aims for this model: To implement an external CV constraint and to address the shortcomings of the CQS model. The model has in the main performed well on both counts. Studies 4, 5 and 6 show that the external constraint system works as expected, that the biasing of the output competition leads to preservation of CV status in errors, and that the additional complexity required to implement external constraints has a payoff in reducing the error rate. As with GBD patients the preservation of CV status is not absolute and the preservation rate may be varied by adjusting the level of the CV bias parameter. The model has also successfully improved on the performance of the CQS model in several areas, as follows:

1. **Word length effects.** Study 1, Figure 6.3, shows a close fit between the model and patient AS for performance vs. word length, for all but the longest words (8-letters). Figure 6.4 shows that this fit is somewhat better than that of the CQS model.

2. **Serial position effects.** The CQX model gives a better fit to the patient data for serial error incidence curves. The peak in end of word positions for CQS is removed and the individual curves for different error types are similar to those produced by patients.

3. **Error types.** The CQS model shows a high incidence of substitution errors involving the re-activation of the first letter in the last-but-one position. This is not evident in the new model, which produces a range of different errors.

4. **Error proportions.** The relative proportions of different error types are considerably less affected by parameter changes in the CQX model than the CQS model. However, two features of the model's performance suggest that the mechanisms within the model for deletion and insertion errors may not be appropriate for the spelling domain. Firstly, the manipulation which must be applied to the letter node activation function (equation 6.6) in order to achieve realistic levels of deletion errors is incompatible with the production of realistic levels of insertions, and vice-versa. Secondly, the model shows a trend of decreasing proportions of deletions with increasing word length, which is incompatible with the

prevailing trend among GBD patients. This problem is considered in more detail below.

5. **Effect of lexical status.** The CQX model produces quantitatively worse but qualitatively very similar performance when the level of noise is slightly increased. This provides a good match to GBD patients, and can be interpreted as a greater vulnerability to noise for nonword stimuli.

6. **Double letters.** The CQS model never deletes or inserts geminate features, in contrast to GBD patients. The CQX model makes the correct types and proportions of geminate errors.

7. **Parameters.** The CQX model is considerably simplified in its basic structure, and as a consequence has many fewer parameters than the CQS model.

6.4.1 Deletion mechanisms

The main residual problem for the model is the detailed structure of the errors, and two specific issues can be identified: Insertion and deletion errors show a tendency to occur with lower frequencies than in GBD patients, and deletion and substitution errors show the incorrect trends with respect to incidence at different word lengths. The mechanisms behind the relative incidences of different types of error in a CQ system are non-linear and far from trivial, and a full exploration of the issues involved is beyond the scope of the current work. However, it is possible to make a specific suggestion regarding an alternative mechanism for deletion errors which has some impact on these problems. First it will be necessary to look in more detail at insertion and deletion errors.

The mechanisms for insertion and deletion errors in a basic CQ system such as CQS or CQX share the feature that they involve a cascade of letter movement errors following an initial erroneous response. In the case of an insertion each letter in this 'cascade region' is produced one position too late. Deletion errors also require a cascade of minor errors, in this case each letter following the initial error appears in the position *before* its target position. The two types of error thus involve letter movements in

opposite directions in the cascade region. The propensity for letters to move in either direction can be influenced by manipulating the model's dynamics, as with the use of the 'gain control' parameter in the CQX model to increase the incidence of deletions. However, increasing the tendency to move in one direction decreases that to move in the other - thus the CQX model makes very few insertion errors.

An additional problem is that the chance of another error occurring are higher in the cascade region than normal, since both the absolute activation level of the most active letter and its winning margin are reduced when the letter is not in its target position. The cascade region is thus more than normally susceptible to errors which may mask the initial insertion or deletion. Clearly this is more likely to happen the longer the vulnerable cascade region is, so the nearer an insertion or deletion is to the end of the word the more likely it is to survive intact. Hence the skewing of the serial incidence curve for 'pure' insertions and deletions towards the ends of words in the CQS model.

Taken together, these two problems suggest that the 'cascade of errors' may not be an appropriate mechanism for insertions and deletions in GBD. There do appear to be plausible alternative mechanisms for both error types which could be explored by future models. A possible mechanism for deletions is the introduction of an activation threshold which must be exceeded before a response can be made by the model (an approach which has already been taken in some CQ models, in particular in Page and Norris's (1997; submitted) Primacy model). A related mechanism can also be suggested for insertion errors: It might be made possible for two letters to be produced within a single time-step if they were very close together in activation.

It is possible to test the suggestion with relation to deletion errors without making major changes to the CQX model. Table 6.7 shows the effect of word length on error proportions for a version of the model with an activation threshold which must be exceeded by the winning letter node on a particular time-step for output to be generated for that time-step. The threshold is set to 0.6, and the letter node gain control parameter, g, is set to 0, since the asymmetry introduced by this parameter is no longer required for the model to produce a reasonable number of deletion errors. All other

parameters are as shown in Table 6.1, apart from the noise level which is set to 0.55 in order to give a comparable level of performance (52% correct for 6-letter words).

Word length	Insertions	Deletions	Exchanges	Shifts	Substitutions
3 letters	11.1%	7.8%	11.1%	0%	70.1%
6 letters	3.9%	14.2%	19.5%	1.1%	61.3%
8 letters	2.8%	7.5%	22.6%	2.0%	65.1%

Table 6.7. *Effect of word length on error proportions for a modified version of the CQX model with an activation threshold on letter nodes. Figures shown are percentages of single-error responses.*

Although the modification applies directly only to deletion errors it affects the trend in deletion, exchange and substitution errors. Between 3 and 6-letter words the modified model now shows the correct qualitative trend - an increase in deletions and a decrease in substitutions with increasing word length. Between 6 and 8-letter words, however, these trends are reversed. Clearly this modification on its own will not fully address the issue. However, the example does demonstrate that trends in the error pattern are affected by relatively simple changes to the operation of the model, and an apparently incorrect result at this level of detail need not rule out the general approach. Further work will be required to elucidate the implications for the sequencing system of the apparently stable trends in the GBD patient data.

In summary, the CQX model provides a more comprehensive explanation of the GBD data and a closer fit than does the CQS model. It has been argued that much of this improvement is due to the increase in positional discrimination afforded by the RBF activation function. The good performance of the model, and the stability of the 'core' features of the error pattern, provides some reassurance that the underlying CQ sequencing framework is the correct way to address the output stages of the spelling process.

7

CQM: Towards a multi-layer network CQ model

This chapter will take a rather different approach to the problem of representing consonant/vowel status in a model of spelling, and in the process a novel network architecture will be developed. While the previous chapter refined the initial CQS model, adding an external constraint system to model the effect of consonant/vowel letter status, this chapter takes as its starting point the idea that the same effect might be explained by regularities in an internal representation - the internal constraint approach of Chapter 5. It would be possible to 'manufacture' a representation for letters which explicitly separated consonants from vowels, in the same way that articulatory loop models have used a distributed phonemic representation to explain acoustic similarity effects. However, the existence of internal representations suggests an altogether more interesting approach to the problem - the use of a multi-layer connectionist model.

Connectionist models which have employed the CQ approach to sequence generation have all been of the 'single layer' type - that is, only a single layer of modifiable weights is employed. Where models include the ability to learn, a single layer arrangement allows learning algorithms of the Hebbian type to be used, where the weight on a connection is modified according to the degree of co-activation between the source and destination units for the connection. This type of learning rule is more physiologically plausible than those which must be employed in networks with multiple layers of modifiable weights and 'hidden' units, such as those of the 'backpropagation' type. However, multiple layer networks (often known as multi-layer perceptrons, or MLPs, when units use simple 'dot product' activation arrangements) are considerably more powerful than their single-layer cousins. Might an MLP CQ model be possible, and if so could it offer any advantages over single-layer models?

Several limitations of the fully localist, two-layer architecture of previous models prompt the investigation of variations based on multi-layer networks. For example:

- If multiple different sequences are to be stored in memory, a separate pair or set of 'timing' or 'context' nodes (depending on the model) must be created to represent each sequence.

- No efficiencies of generalisation are possible in the storage of several very similar sequences. Each must be stored separately.

- The generation of timing information which drives sequential recall in the CQ paradigm is integral to the representation of the sequence. Again, there are no opportunities for generalisation across sequences based on temporal position.

- The problem of combining information sources, as when phonological and lexical information are combined in a two-route spelling model, is difficult (see Glasspool, Houghton and Shallice, 1995). An MLP model with multiple input fields may provide an interesting basis for starting to tackle this problem.

- Hitherto, it has only been possible to make general comments about the suitability of unstructured noise as a model for lesion damage in CQ models. With a model which is substantially distributed in its representations it will be possible to use true disconnection lesions as well as noise disruption.

7.1 Aims

As this model will use a novel architecture a major aim will be to establish its effectiveness both as a generator of serial behaviour and as a model for the general pattern of serial errors identified in Chapter 1, in particular:

- The typical dynamics of CQ sequence production, with pre-activation of responses.

- The effect of sequence length on performance.

- Bowed serial error curves.

- The occurrence of order errors, including exchanges.

A further general aim will of course be to provide a specific model for GBD. The intention will be to provide a qualitative rather than quantitative match to the data. Finally, since this approach to modelling has been prompted by the idea of using internal constraints to model the preservation of CV status in GBD errors, the generation and expression of such constraints will be a target.

The modelling of geminate errors is not an aim. The models of previous chapters have demonstrated the general effectiveness of a separate geminate representation in qualitatively modelling these phenomena. The arguments supporting this approach hold good for this model, but the inclusion of a geminate system would add unnecessary complexity and will be left for future work. As with the models of previous chapters, reproducing exactly the pattern of relative incidence of different error types will also be considered of less importance than the more general aims stated above, as a detailed investigation of the mechanisms behind different types of error is beyond the scope of the present work.

7.2 Architecture and operation

Figure 7.1 shows the architecture of the model, CQM[7]. Since the CQ approach depends on the localised inhibition of a single output item at each time step in sequence production, a set of localist item nodes are used at the output layer. CQ dynamics require the selection and then temporary inhibition of the most active item at each time step, represented in the diagram by a notional competitive filter. The output nodes are activated via a hidden layer of nodes, which in turn receive input from two fields of input nodes, one representing the identity of the sequence (in this case, word) to be produced and the other the current position in the sequence. Other than the one-to-one connections between output nodes and the competitive filter, all connections in the network are modifiable and may be set by a backpropagation training procedure, described below. The task on which the network is trained is that of generating the correct sequence of letters at the output given a particular word identity representation

---

[7] For CQ in an MLP model

(which is held steady during sequence production) and sequence position information (which runs through a standard sequence of states during sequence production). The additional output nodes labelled C and V are explained below.
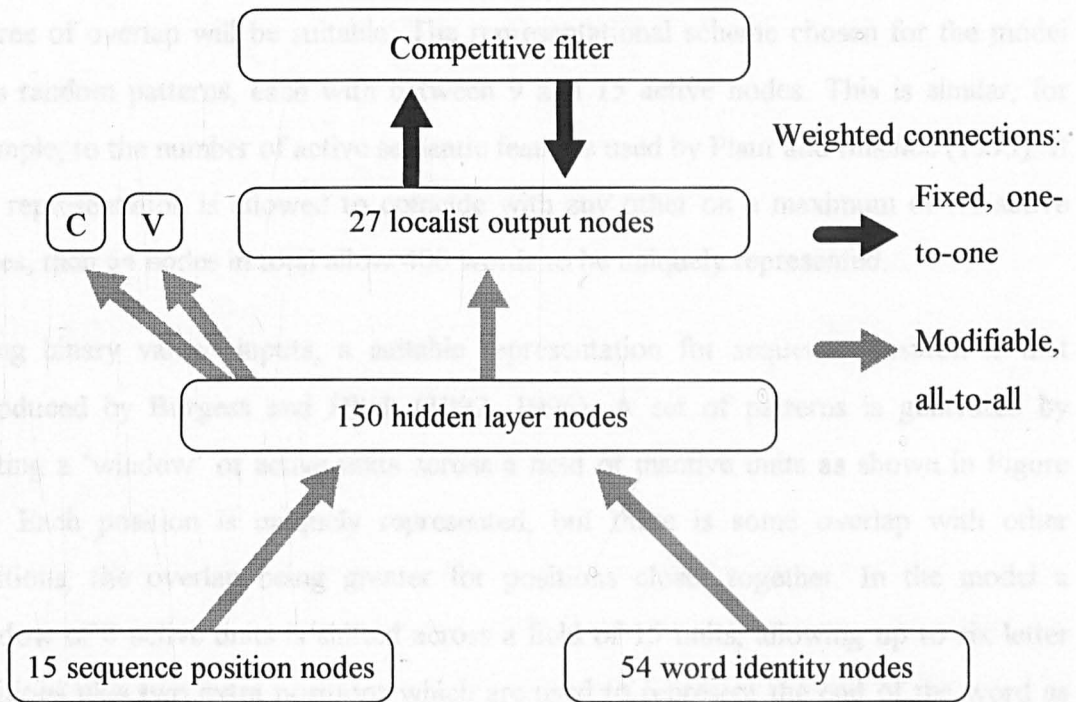
Figure 7.1. *The architecture of the CQX model. See text for details.*

7.2.1 Representations

While learning rules of the backpropagation type are capable of learning mappings between representations which use arbitrary activation levels, it is common to use binary-valued representations with values of +1 and 0 (or +1 and -1) in backpropagation networks. For convenience this convention will be followed here.

Input Representations

Taking the word identity representation first, this would correspond, on a general spelling framework, to an entry in either the output graphemic lexicon or the semantic system, depending on how far down the 'lexical route' the model is held to be situated.

It is not entirely clear what information this representation might include, but its semantic source suggests that there should be some similarity in the representations for semantically similar words. Since an accurate semantic representation is not required for the purposes of this model, and there is unlikely to be any systematic relationship between semantic content and spelling, unique randomly chosen vectors having some degree of overlap will be suitable. The representational scheme chosen for the model uses random patterns, each with between 9 and 15 active nodes. This is similar, for example, to the number of active semantic features used by Plaut and Shallice (1993). If any representation is allowed to coincide with any other on a maximum of six active nodes, then 54 nodes in total allow 400 words to be uniquely represented.

Using binary valued inputs, a suitable representation for sequence position is that introduced by Burgess and Hitch (1992, 1996). A set of patterns is generated by shifting a 'window' of active units across a field of inactive units as shown in Figure 7.2. Each position is uniquely represented, but there is some overlap with other positions, the overlap being greater for positions closer together. In the model a window of 8 active units is shifted across a field of 15 units, allowing up to six letter positions plus two extra positions which are used to represent the end of the word as detailed below.

Output representations

The model uses a localist output representation: A set of 26 nodes represent the letters of the alphabet. Stopping spelling is a concern for this model: The preferred method would be to use an activation threshold to stop sequencing as in Chapter 6. However, the current model will be investigated using lesions at various different points, and it is not therefore possible to confine noise disruption to the competitive filter. Consequently the 'stop symbol' technique of Chapter 4 is used. An additional node in the output letter field, treated in the same way as letter nodes, represents the 'end of word' marker which causes spelling production to stop. A (simulated) competitive filter selects the most active letter node at each time step, and thereafter inhibits it by setting its activation level to a uniform negative level from which it recovers slowly.

Time step

```
1  ● ● ● ● ● ● ● ● ○ ○ ○ ○ ○ ○ ○
2  ○ ● ● ● ● ● ● ● ● ○ ○ ○ ○ ○ ○
3  ○ ○ ● ● ● ● ● ● ● ● ○ ○ ○ ○ ○
4  ○ ○ ○ ● ● ● ● ● ● ● ● ○ ○ ○ ○       ● Active node
5  ○ ○ ○ ○ ● ● ● ● ● ● ● ● ○ ○ ○
6  ○ ○ ○ ○ ○ ● ● ● ● ● ● ● ● ○ ○       ○ Inactive node
7  ○ ○ ○ ○ ○ ○ ● ● ● ● ● ● ● ● ○
8  ○ ○ ○ ○ ○ ○ ○ ● ● ● ● ● ● ● ●

   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
```

'Sequence position' nodes
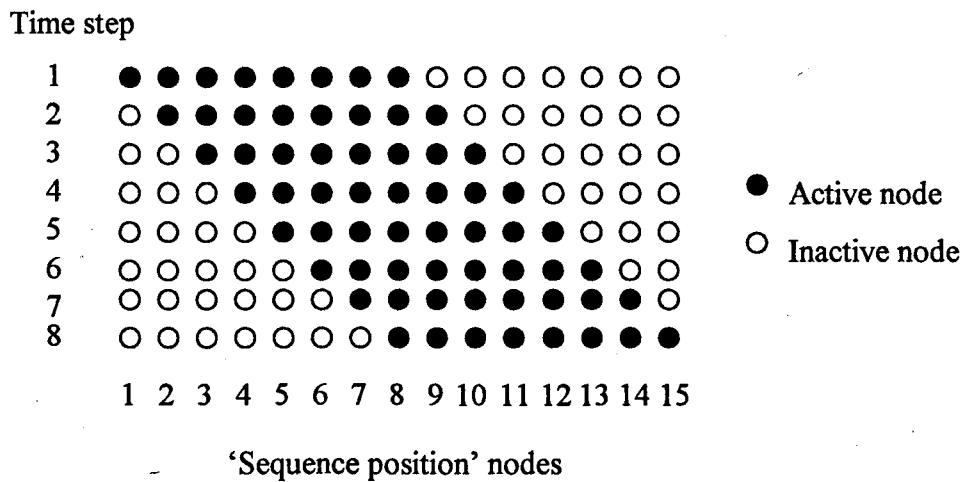
Figure 7.2. *Representation of sequence position. A 'window' of 8 active nodes is shifted across a field of 15 inactive nodes, yielding vectors to represent eight spelling time steps.*

Hidden layer

With a training set of 400 words a hidden layer of 100 nodes is insufficient for the model to learn to spell all the words accurately. 150 nodes support correct learning, but increasing the layer to 200 nodes gives no improvement in the behaviour of the net. 150 hidden nodes were therefore used in all the simulations described below.

7.2.2 Training procedure

Conventionally, backpropagation algorithms (Rumelhart, Hinton and Williams, 1986) operate by feeding the error at each output unit for each target pattern back to modify the weights responsible for it, the assumption being that the activation level of every output unit is equally important for the generation of a correct output. For this model the requirement may be relaxed considerably, as all that is required at each time step is that the correct letter be the most active. The actual activation levels involved are not important (although it is useful to generate a fairly uniform high activation level on each 'winning' letter to provide good immunity from noise), neither are the relative activation levels of all but the winning letter. A 'lazy' learning rule is thus used. The error for each letter node is calculated at each time step as follows:

- The error for the letter which should win – the target letter – is the difference between its activation and 1.0.

- For all other nodes, no error is fed back at all unless they are more active than the target node. In this case their error is calculated by comparison with 0.0.

The error values thus calculated are then used in a standard cross-entropy backpropagation learning algorithm (Hinton, 1989)

A similar training regime is operated to that of the CQS model of Chapter 4. Firstly, a small margin, ϕ, is applied to the activation comparisons in the competitive filter to improve the robustness of recall. (The margin is subtracted from the activation of each node before it is compared with others to find the 'winner'). Secondly, the learning process performs repair during production: After each time step during learning the *target* letter is inhibited, regardless of whether the target actually won the output competition.

CV status

In order for the model to produce a CV preservation effect in errors due to internal representational regularities, it clearly needs to learn the distinction between consonants and vowels. One possibility is that the network might learn this distinction simply by exposure to the set of words in its training corpus. In order for this to occur, the C/V distinction would have to represent a useful regularity in the input to output mapping which the learning rule could exploit in order to simplify the mapping task. However, the consonant-vowel distinction is primarily a phonological one and is not likely to be of utility in mapping from abstract word identity or sequence position to letter identity. It is reasonable to presume though that information on CV status is available from phonology, and such information may be used as part of the output representation. In order to achieve this, an extra pair of nodes are added at the output layer to indicate the CV status of each letter. During training the network is required to activate the C or V node in parallel with the appropriate letter node. The state of the CV nodes is ignored during testing. (This has some similarities with the 'hints' approach of Christiansen, 1997, though the motivation is different).

7.2.3 Formal description

The operation of the network can be separated into two passes. During the forward pass, an activation pattern is applied to the input layer and activation propagates forward to the output layer. The backward pass operates only during training, when an error signal is propagated back from the output layer to adjust weights in proportion to their contribution to the error.

Forward pass

Nodes in the input fields have their activation levels set to 1.0 or 0.0 according to the current input pattern. The net input $net_i(t)$ to node i in the hidden or output layer at time step t is given by:

$$net_i = \sum_{j=1}^{n} A_j W_{ji}$$

Equation 7.1

where A_j is the activation of node j in the previous layer, and W_{ji} is the weight from node j to node i.

The activity $A_i(t)$ of node i in the hidden layer at time step t is given by:

$$A_i(t) = f(net_i(t))$$

Equation 7.2

The function f is the logistic function standardly used in backpropagation networks (Rumelhart, Hinton and Williams, 1986):

$$f(x) = \frac{1}{e^{-x} + 1}$$

Equation 7.3

As well as receiving input from nodes in the previous layer, each node in the hidden and output layers also receives a bias input which may be thought of as an additional weight from a unit which is permanently set to an activation of 1.0.

The C and V nodes in the output layer also obey equation 7.2. However, as in the previous models, letter nodes recover slowly from inhibition. The activation $A_i(t)$ of letter node i at time t is given by:

$$A_i(t) = \begin{cases} net_i(t) & \text{if } A_i(t-1) \geq 0 \\ net_i(t) + rA_i(t-1) & \text{otherwise} \end{cases} \qquad \text{Equation 7.4}$$

Where r is a parameter which governs the rate of recovery from inhibition.

The most active letter node is determined by a simulated competitive filter, and the activation level of this letter node is then set to a standard negative (inhibited) activation level, *Inh.*

Backward pass

As Hinton (1989) shows, in a network like the current one where binary valued output vectors are desired, and real-valued output vectors may be interpreted as probability distributions over binary vectors (the CQ noisy selection procedure is straightforwardly interpretable in this way), the appropriate error measure to use in a backpropagation training procedure is the cross-entropy, C, between the desired and actual probability distributions, rather than the more usual sum-squared-error (SSE) measure. The cross-entropy between an actual probability vector A, with elements a_i, and a desired probability vector D with elements d_i, is given by:

$$C = -\sum_i d_i \log_2(a_i) + (1 - d_i)\log_2(1 - a_i) \qquad \text{Equation 7.5}$$

When used in a backpropagation procedure the derivative of C is multiplied by the derivative of the logistic function, and an advantage of using the cross entropy function is that it then reduces simply to the difference between the desired and actual outputs (Hinton, 1989). An error value δ_u is thus generated for each output layer unit as follows:

$$\delta_u = (d_u - a_u) \qquad \text{Equation 7.6}$$

where d_u is the desired activation value for output unit u and a_u is the actual value. The desired values d_u are generated according to the 'lazy' learning rule detailed above. A

weight change is calculated for the weights from the hidden layer to the output layer according to:

$$\Delta W_{uh} = \varepsilon \delta_u A_h$$

Equation 7.7

where ΔW_{uh} is the required change in the weight from hidden unit h to output unit u, A_h is the activation of hidden unit h and ε is a small constant, the 'learning rate'. The cross-entropy error function requires a lower value for ε than is usual with standard backpropagation networks using the SSE error function. A new error value is derived for each hidden unit by propagating the output error values back along the weights. Here the values must be multiplied by the derivative of the logistic function, so δ_h, the error value for hidden unit h, is given by:

$$\delta_h = A_h\left(1 - A_h\right)\sum_u \delta_u W_{uh}$$

Equation 7.8

A weight change is now calculated for the input - to - hidden layer weights using this new error value:

$$\Delta W_{hi} = \varepsilon \delta_h A_i$$

Equation 7.9

where ΔW_{hi} is the required change in the weight from input unit i to hidden unit h, and A_i is the activation of input unit i.

For each weight W in the network, the weight changes calculated above are now applied using:

$$W(t) = W(t-1) + \Delta W(t) + m \, \Delta W(t-1)$$

Equation 7.10

where $W(t)$ is the new weight value for time step t, $W(t-1)$ is the value of the weight at the previous time step, $\Delta W(t)$ is the required weight change for the current time step, $\Delta W(t-1)$ is the required weight change calculated on the previous time step, and m is a small ($0 < m < 1$) momentum value. At the start of learning, all weights $W(t)$ are set to random values between ± 0.5, and all $\Delta W(t-1)$ are assumed to be 0.

Table 7.1 gives the parameter values used in the simulations of the following section.

Parameter	Symbol	Value
Letter node recovery rate	r	0.8
Letter node inhibition level	*Inh*	-1.5
Momentum	m	0.9
Learning rate	ε	5.0×10^{-4}
Learning margin	ϕ	0.1

Table 7.1. *Parameter values used in simulations.*

7.3 Training

7.3.1 Training set

The model was trained on a set of 400 monosyllabic words selected from the MRC Psycholinguistic Database[8], 100 each of length 3, 4 5 and 6 letters. The 100 most frequent (on the index of Kucera and Francis, 1967) monosyllabic words with no repeated letters were selected from the database for each word length. In each training epoch every word was presented to the network once, in random order.

7.3.2 Performance

Figure 7.3 shows the SSE (which is a useful measure of error for visualisation purposes although not the one used by the training procedure) and the percentages of words and letters correctly spelled at each epoch (complete presentation of the corpus) during the training of the network.

---

[8] From the Oxford Text Archive, and on the World Wide Web at:
http://www.psych.nwu.edu/psych/people/resappt/yamada/dict.html
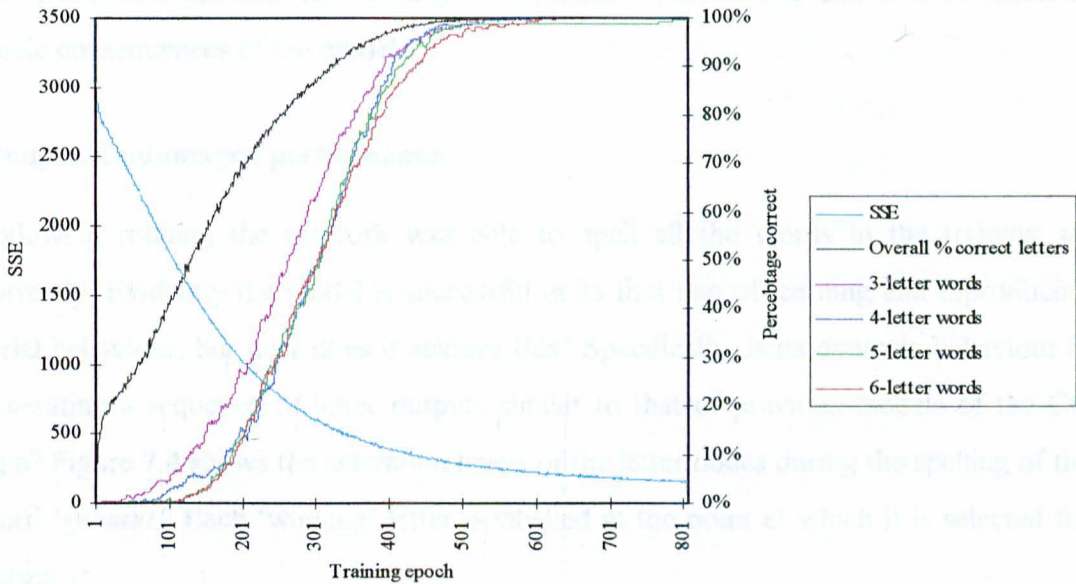
Figure 7.3. *Learning to spell: The sum-squared-error (SSE) and percentage of letters and words which could be correctly spelled after each epoch in training the network.*

The SSE measure uses the delta values developed at the output layer by the 'lazy' learning rule and thus gives an indication of the progress of learning, but the learning rule is less straightforwardly interpreted as minimising this function than usual backpropagation rules due to the non-linear nature of the competitive mechanism. The number of correct letters is perhaps a more accurate measure of progress. The network learns three-letter words more quickly than longer words, and first correctly spelled all of the words in the corpus after 769 training epochs. Training was continued after this point until twenty epochs with all words spelled correctly were achieved, at 807 epochs, to allow the network to stabilise (the performance fluctuates slightly from epoch to epoch as the learning rule still generates error values for correct letters while they are less than maximally activated).

7.4 Simulations

7.4.1 Stability of network

The studies reported below were repeated twice from different random starting weights with qualitatively very similar results. A version of the network with 170 hidden units

also performed qualitatively similarly. The results reported here can thus be taken as stable consequences of the model.

Study 1. Undamaged performance

Following training the network was able to spell all the words in the training set correctly. Evidently the model is successful in its first aim of learning and reproducing serial behaviour, but how does it achieve this? Specifically, is its dynamic behaviour in generating a sequence of letter outputs similar to that of previous models of the CQ type? Figure 7.4 shows the activation levels of the letter nodes during the spelling of the word *'ground'*. Each 'winning' letter is labelled at the point at which it is selected for output.
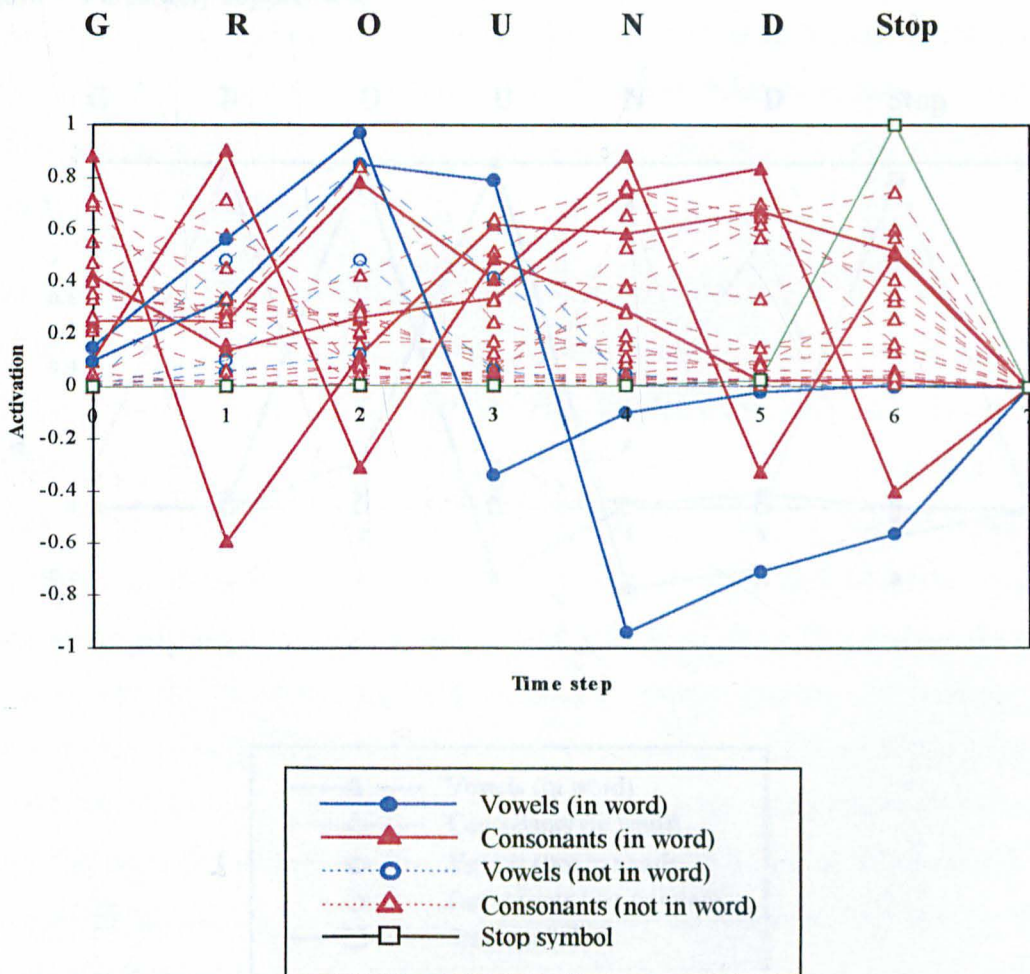
Figure 7.4. *The dynamics of the recall process. Activation levels of output letter nodes at each time-step during the recall of the word 'ground'.*

The model shows the type of dynamic behaviour which is typical of previous CQ models - the general pattern is for letters to become active before they are selected, increasing in activation as their target position in the sequence approaches until they exceed other letter activations and are selected for output. The pre-activation of letters is not something the network is trained to do - the 'lazy' learning rule makes no demands on any letters which are not the target in the current position, except that they should have a lower activation level than the target letter itself.

It is interesting to compare Figure 7.4 with Figure 7.5, which shows the output dynamics as the same word is spelled by a version of the model trained without using the 'lazy' learning procedure. The error values in equation 7.6 are here developed for all output items at every time step. The result is that the pre-activation of items seen in Figure 7.4 is largely suppressed.
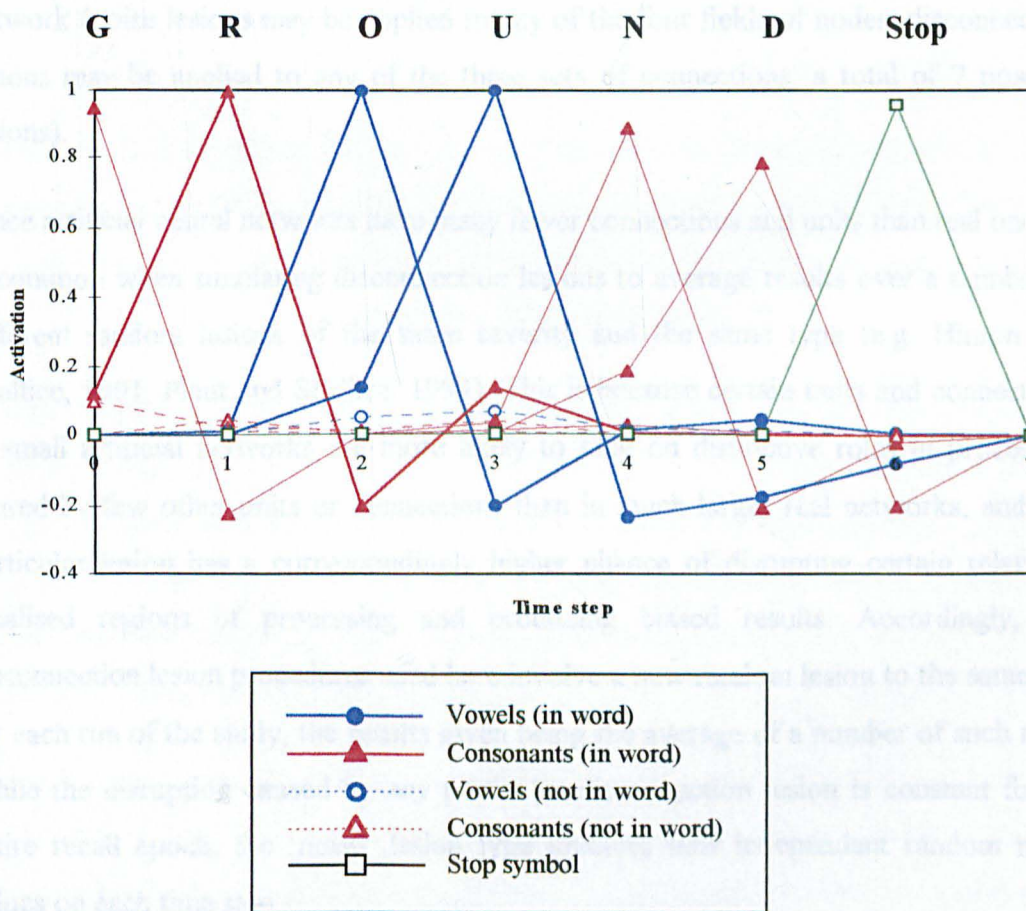
Figure 7.5. *Letter node dynamics while spelling the six-letter word 'ground', from a version of the CQX model trained without using the 'lazy' learning rule.*

It is clear from Figure 7.4 that many letters which are not present in the target word acquire a high level of activation during the spelling process (especially by comparison with similar graphs for the localist models of Chapters 4 and 6). This is due to the overlap between the input patterns (on both the word identity and sequence position fields) of the target word and those of other words the model has learned.

Study 2. Lesioning the network

A number of procedures are possible by which the performance of the trained network may be disrupted. Two lesioning procedures are used here - the addition of random noise to the activations of nodes at any layer ('noise lesions'), or the removal of a proportion of the connections between two layers ('disconnection lesions'). The latter procedure is of course a new option not available on the localist models of Chapters 4 and 6. This section examines the effect of both lesion types at each possible point in the network (noise lesions may be applied to any of the four fields of nodes, disconnection lesions may be applied to any of the three sets of connections, a total of 7 possible lesions).

Since artificial neural networks have many fewer connections and units than real ones, it is common when simulating disconnection lesions to average results over a number of different random lesions of the same severity and the same type (e.g. Hinton and Shallice, 1991, Plaut and Shallice, 1993). This is because certain units and connections in small artificial networks are more likely to take on distinctive roles in processing shared by few other units or connections than in much larger real networks, and any particular lesion has a correspondingly higher chance of disrupting certain relatively localised regions of processing and producing biased results. Accordingly, the disconnection lesion procedures used here involve a new random lesion to the same site for each run of the study, the results given being the average of a number of such runs. While the disruption caused by any particular disconnection lesion is constant for an entire recall epoch, the 'noise' lesion type involves new independent random noise values on each time step.

Figure 7.6 shows the result of lesions of each type, at each possible site on the network, on the overall performance of the network spelling words of length three, four, five and

six letters. For parity with the CQX model (and patient AS) the severity of each lesion is again normalised to give a standard performance level of approximately 55% correct for six-letter words.
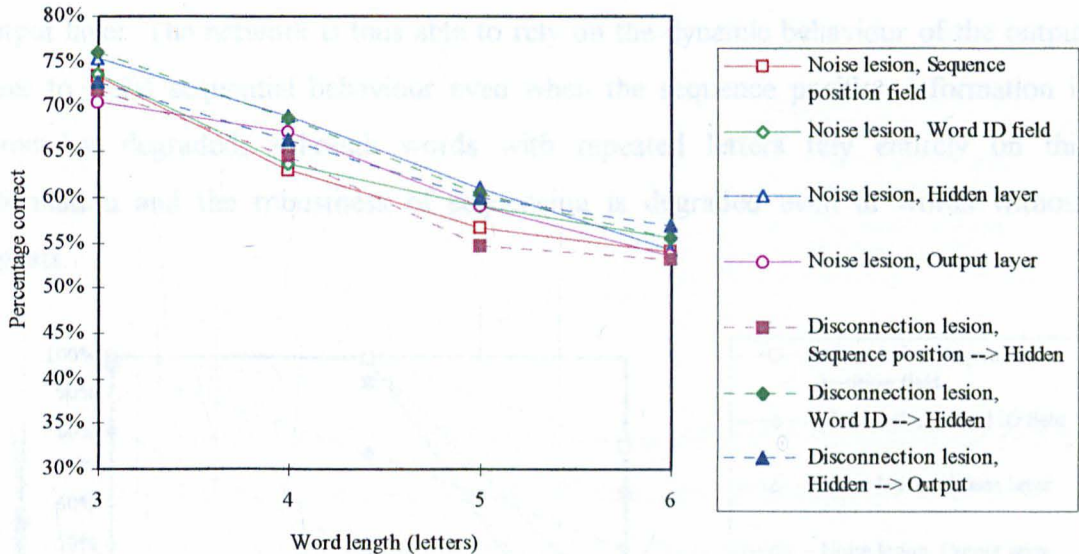
Figure 7.6. *The effect of different lesion types and sites on the performance of the network. The severity of each lesion is set to give approximately the same level of performance on six-letter words. All results in this and subsequent figures are averaged over 500 recall epochs.*

In each case the network shows a clear effect of word length. The different lesion types all produce lines with similar gradients. Qualitatively, this effect is in line with a number of types of serial behaviour where performance decreases as sequence length increases (e.g. speech and verbal STM) as well as GBD. The effect is, however, somewhat milder than that shown by those GBD patients with more severe deficits. Possible reasons for this are considered in section 7.5.

Figure 7.7 shows the effect of lesion severity for each of the lesion types and sites on the recall of 6-letter words. Lesions at later stages in the network - to the hidden or output layers or the connections between them - have a greater effect on performance than lesions nearer the input. Since both input fields contribute to the input pattern of the network at any time step, it is to be expected that lesions to one or other field will

have a lower relative effect on overall performance than a lesion to the whole of the hidden or output layer. The model is least sensitive to damage to sequence position information. This is consistent with the fact that the sequence position field shares the task of providing the impetus for sequencing with the dynamic behaviour built in to the output layer. The network is thus able to rely on the dynamic behaviour of the output layer to assist sequential behaviour even when the sequence position information is somewhat degraded, although words with repeated letters rely entirely on this information and the robustness of sequencing is degraded even in words without repeats.
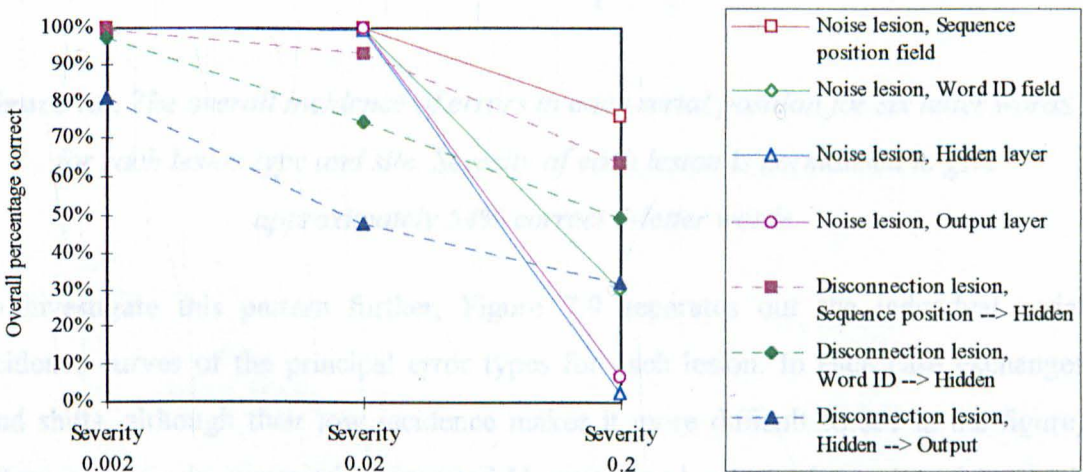
Figure 7.7. *The effect of lesion severity for each of the lesion types and sites on the recall of 6-letter words.*

Looking now in more detail at the patterns of errors produced by lesions, Figure 7.8 shows the overall incidence of errors in each serial position for six letter words, for each lesion, with lesion severity again normalised to approximately 55%. The lesions fall into two classes: Those which occur early in the network - noise lesions on the two input fields or disconnection lesions between the input layer and the hidden layer - result in an increase in error incidence across each serial position to a peak in the final position. Lesions later in the network - noise lesions to the hidden or output layers or disconnection lesions between them - produce profiles showing mild recency effects.
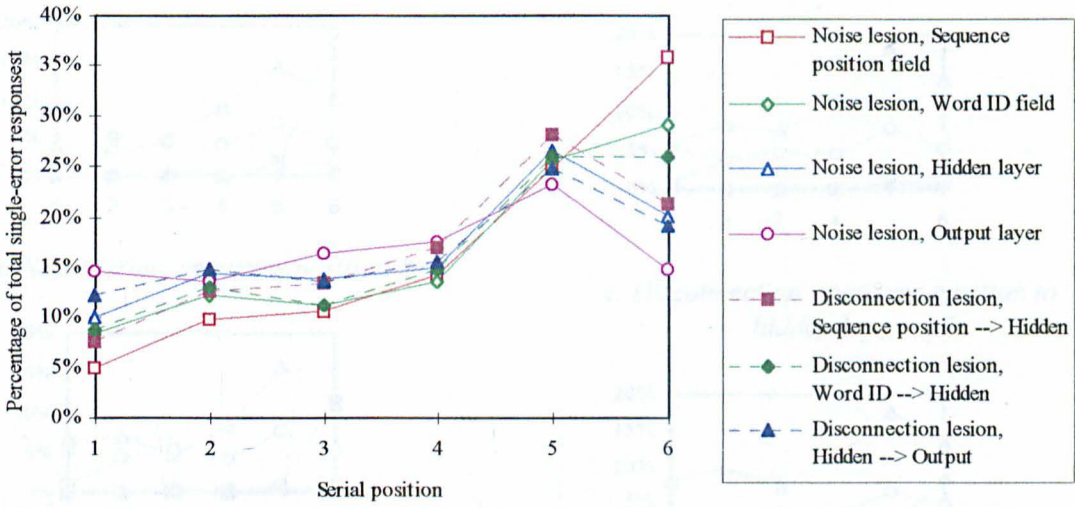
Figure 7.8. *The overall incidence of errors in each serial position for six letter words, for each lesion type and site. Severity of each lesion is normalised to give approximately 54% correct 6-letter words.*

To investigate this pattern further, Figure 7.9 separates out the individual serial incidence curves of the principal error types for each lesion. In each case exchanges (and shifts, although their low incidence makes it more difficult to see in the figure) follow a relatively symmetrical inverted-U curve, and substitutions also show both primacy and recency effects. Insertions and deletions peak in final positions. The main difference between lesions is the incidence of deletions, and this is the cause of the differences in overall profile of Figure 7.8.

Clearly the incidence of deletions, at least, varies with lesion site. Table 7.2 enables a more accurate comparison of the serial incidence of each of the five basic errors for each lesion, again for 6-letter words with lesion severity normalised to give 54% correct. Exchange errors occur with each lesion type and are generally frequent. This is a confirmation that the Competitive Queuing process is operating successfully during sequence generation. The generally rather low rate of insertion and deletion errors is a concern with respect to modelling GBD, with deletions in particular being an important feature of both GBD and normal speller's error profiles. This issue will be returned to in the discussion section.

a. Noise lesion, sequence position field.

b. Noise lesion, word identity field.
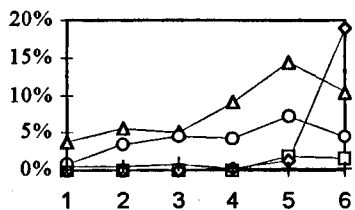
c. Noise lesion, hidden layer.

d. Noise lesion, output layer.
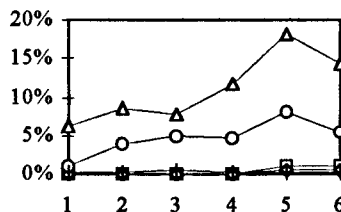
e. Disconnection, sequence position to hidden layer.
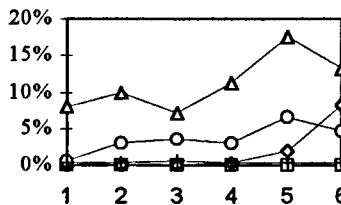
f. Disconnection, word id to hidden layer.
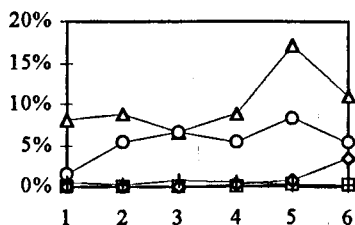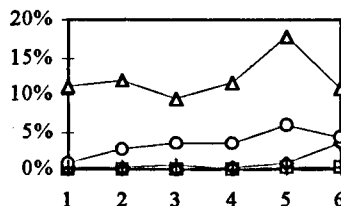
g. Disconnection, hidden layer to output.

—□—	Insertion
—◇—	Deletion
—△—	Substitution
—○—	Exchange
—+—	Shift
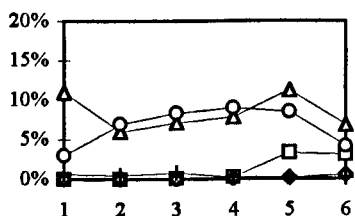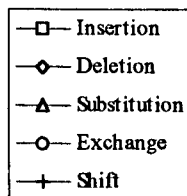
Figure 7.9. *Serial incidence curves for the basic error types following each lesion. Severity of each lesion is normalised to 54% performance level.*

	Insertion	Deletion	Substitution	Exchange	Shift
Noise lesion Sequence position field	4.2%	23.8%	55.8%	14.6%	1.6%
Noise lesion Word ID field	1.4%	15.7%	64.7%	16.9%	1.3%
Noise lesion Hidden layer	0.3%	5.3%	73.0%	19.7%	1.7%
Noise lesion Output layer	8.5%	1.1%	63.4%	25.3%	1.7%
Disconnection lesion Sequence position → Hidden	2.2%	1.2%	78.7%	16.6%	1.3%
Disconnection lesion Word ID → Hidden	0.1%	11.7%	75.3%	12.1%	0.9%
Disconnection lesion Hidden → Output	0.4%	5.2%	81.6%	11.9%	0.8%
Average	2.4%	9.1%	70.5%	16.7%	1.3%

Table 7.2. *Errors for each lesion broken down into the five basic error types. All results for 6-letter words with lesion severity normalised to give approximately 54% correct 6-letter words.*

Finally, Table 7.3 shows the effect of word length on relative proportions of different error types, averaged over all seven lesion types. As with Table 6.3, this data must be treated with care, and again an exploration of the detailed mechanisms behind the error

patterns is beyond the scope of this work. However, encouragingly, the correct trends are shown for deletion and substitution errors with respect to the majority of GBD patients - an increase and decline, respectively, with increasing word length.

Word length	Insertions	Deletions	Exchanges	Shifts	Substitutions
3 letters	10.9%	2.9%	5.5%	0%	80.2%
6 letters	2.4%	9.1%	16.7%	1.3%	70.4%

Table 7.3. *Effect of word length on error proportions. Figures shown are percentages of single-error responses averaged over all seven lesion types, with lesion severity normalised to give approximately 54% correct 6-letter words.*

Study 3. The effect of CV information.

Figure 7.4 shows some co-activation of consonants with consonants and vowels with vowels, although some consonants are also activated in vowel positions. It is not easy to tell from the figure alone to what degree an internal CV constraint has been established. Table 7.4 shows the effect of CV information on recall by comparing the CV preservation rate in transposition and exchange errors with those of an identical version of the model trained without the C and V output layer nodes. In the table 'late' lesions are shown separately - these again are noise lesions to the hidden and output layers and the disconnection lesion between them.

The incidence of vowels in writing is smaller than the incidence of consonants, so it is not surprising that the level of CV status preservation is higher for consonants than for vowels in every case. However, two main points are evident from Table 7.4. Firstly, the level of CV preservation is clearly higher for all lesions in the version of the model which includes specific training on CV status (in the form of C and V nodes at the output layer). Secondly, the in the 'with CV' model, 'late' lesions show a higher level of CV preservation than 'early' ones, presumably indicating that it is in the hidden layer representation that the CV status distinction is made. The 'without CV' model shows no difference between 'early' and 'late' lesions in this respect. The C and V nodes at

the output layer are necessary for the network to form internal representations which distinguish C and V status strongly.

		Transpositions		Substitutions	
Version	Lesions	C	V	C	V
With	All	75.3%	58.2%	81.5%	49.4%
CV	Late	80.7%	64.5%	85.4%	57.1%
Without	All	63.2%	40.5%	74.4%	27.1%
CV	Late	64.5%	39.2%	75.4%	28.1%

Table 7.4. *Percentage of errors on consonants and vowels which preserve CV status, compared for all lesions or late lesions only, and for nets trained with and without CV nodes in the output layer.*

7.5 Discussion

The main aim for this chapter was to demonstrate that the CQM model could learn the mapping from word identity and sequence position inputs to the correct production of sequential spelling output. In this the model was successful. The process of sequence generation in the model shows the type of dynamic behaviour common to other CQ models with pre-activation of up-coming responses. The overlap in input representation between different words in the vocabulary of the model leads to a greater degree of interference from letters not present in the target word than is typical of previous CQ models, however. The second aim was that the model should exhibit the 'standard' error patterns associated with CQ systems when lesioned. For all lesion types investigated the model shows a clear and reliable effect of word length, produces ordering errors as well as 'item' errors, including exchanges, and for later lesion sites it shows a bowed serial error incidence curve. The third general aim was to demonstrate the ability to acquire internally mediated serial constraints, which the model shows in its

preferential preservation of CV status in errors. Overall, then, the model successfully demonstrates the operation of CQ dynamics and internally generated serial constraints in a multi-layer network. As far as modelling GBD is concerned, the preservation of CV status and the production of the correct trends in deletion and substitution errors with increasing word length are encouraging. However, the model has some shortcomings. While this preliminary model was not intended to capture the full GBD error pattern, it will be useful to consider these in some detail.

Firstly, the effect of word length on performance is less marked than that in those GBD patients with more severe deficits. Patient AS, for example, whose performance on 6-letter words the lesions have been scaled to match, performs at approximately 90% on 3-letter words compared with a performance of 70%-75% by the model. There are several reasons why the model might be less severely effected by word length than patients. The model uses a learning procedure, error backpropagation, with considerably more power than the simple Hebbian weight-setting arrangements of the previous models. Backpropagation is able to direct its resources towards those items in the training set which particularly need attention, and as a result it is possible that longer words, which are more difficult to learn, are given more attention than the relatively easier short words. This may endow them with more robust representations than would otherwise be the case. A second possibility concerns the representation used for sequence position. All words, long and short, are equally likely to share active nodes in the word identity field. In the sequence position field, however, long words share fewer active nodes than short words in proportion to their length, since all words use the first few sequence positions while only longer words use the last few positions. Short words thus have proportionately more potential competitors than long words and this again skews the likelihood of errors in the direction which would explain the flattening of the word length / performance relationship. Finally, the corpus used for training is not particularly naturalistic. The range of word lengths is less than would be encountered naturally, the overall number of words learned is unrealistically small, and the frequency of occurrence of words is not taken into account beyond the selection of the highest frequency words possible. This latter point may be important, since the frequency relationships between word lengths are not preserved in the training set - the

average frequency of the three-letter words is higher than that of the six-letter words, and Zipf's law suggests that this relationship will hold true of language in general. If a training set which reflected these frequency differences more accurately were used (presenting each word for training with a frequency proportional to its natural frequency of occurrence, for example) the expectation would be that shorter words would be better learned by comparison with longer words than they are with the present training corpus, which might lead to a stronger word length effect under lesion damage.

The second shortcoming of the lesioned network as a model for GBD concerns incidence of different error types. For each lesion errors of all types are more common in later letter positions, although substitutions, exchanges and shifts show recency as well as primacy effects. Deletion and insertion errors peak in final letter positions and, except in the case of certain 'early' lesions, are rare by comparison with patients. This is most likely a result of the stopping mechanism chosen for the model. Figure 7.4 shows that the stop symbol has a rather different activation profile to letter nodes, becoming active quite sharply in the correct position. Hence the mechanism for deletions which is inherent in the basic dynamic process of CQ - an initial anticipation error followed by a cascade of further anticipations - is less likely to produce 'clean' deletions since the stop symbol will not easily participate in such a cascade. Evidently some deletions of this kind occur as this is the only mechanism which will explain the occurrence of deletions earlier than the last letter of the word. However, it is on this final letter that the clearest difference in the deletion curve between lesions is evident. Deletions are most common with lesions which disrupt sequence position information, evidently allowing the stop marker to become active early, or word identity information, which allows shorter words to influence the production of longer words and hence has the same effect. These effects may explain the increase in the final peak of the deletion curve with 'early' lesions, but they do not explain the general skew of the deletion curve towards the end of the word. An important factor here may be the greater competition from other letters by comparison with the localist models, which is progressively more likely to interfere with longer 'cascade regions' following a deletion and lead to more complex errors. The greater competition is due to two factors: The considerable overlap between input layer representations for different words (on both

input fields) and the 'lazy' learning rule, which allows competing letters to be activated regardless of their presence in the target word. However, the degree of overlap between the 'word ID' patterns of different words was set somewhat arbitrarily and it would be reasonable to vary this in future work.

As for the CQX model the use of an activation threshold, below which no output is produced, may lead to a better account for deletion errors. However, this is not such a straightforward proposition for the CQM model as for CQX. Simply imposing a threshold on the model during its recall phase is not enough, as the threshold must be present during the training phase too, which requires that the training rule be modified to reinforce 'target' letters only when they fail to exceed the threshold (with the rule as it stands all target letters are under pressure to achieve maximum activation levels). Such changes in the model will be an interesting area for further work.

Noise lesions to either of the input fields result in a relatively large increase in the number of deletion errors, sufficient to skew the overall serial error incidence into a monotonically rising curve. It is interesting to note that GBD patient HR, who produced a large majority of deletion errors, showed a serial error curve of very similar shape. It may prove to be the case that there is more than one mechanism for deletion, in which case the mechanism at work in the current model may better model the deletion mechanism for patients like HR than for those patients who produce relatively fewer and more symmetrically distributed deletions.

Refinements will be required to produce more detailed future models, but the current model has succeeded in demonstrating that the approach is workable and promising as a model of GBD. Further, it has shown that CQ dynamics are possible in an MLP architecture which is potentially applicable to other areas where CQ has been used to model the production of serial behaviour from LTM - in particular, speech production and typing. The development of a more accurate model of GBD using this approach will require a careful investigation of the influence of lexical parameters such as word concreteness and frequency at different points in the network, and will no doubt require a better understanding of the mechanisms of error in CQ models. If these problems can

be overcome this type of model may be an interesting proposition for more complex models of serial processes.

8

Discussion and conclusions

The successes and limitations of each of the three models advanced in this thesis have already been discussed with respect to their own particular goals. This chapter takes a broader look at all three models together, and is organised in three sections of increasing generality. The first summarises the performance and limitations of the models, examining those successes and shortcomings which may reflect on the appropriateness of the underlying CQ approach. The second section compares and contrasts the approaches taken by the CQX and CQM models, looking both at issues raised by the two approaches and at ways in which they might be integrated. The third section discusses more general theoretical issues, focusing on what the models have to say about the spelling system.

8.1 Overview: CQS, CQX and CQM as models for GBD

8.1.1 Summary of model development

As a starting point the CQS model was intended to provide a qualitative account for the gross features of GBD patient spelling errors in terms of the CQ sequencing approach. The model is based on a straightforward implementation of Houghton's (1990) speech production model with minimal changes to allow it to operate in the spelling modality - the representation of letters rather than phonemes by item nodes, and the addition of a geminate mechanism. It demonstrates qualitatively correct effects of word length and serial position on overall error incidence, and the incidence of individual error types shows qualitative similarities to GBD error patterns. The model gives an initial confirmation that the CQ approach is appropriate for modelling spelling.

The CQX model was intended to refine the CQS approach in a number of ways, and thus attempts to provide a more detailed and quantitative account for GBD errors. This model does not simply add external CV constraints to the CQS model, however, but also makes a number of changes to the underlying sequencing system, the most

fundamental being the use of an RBF activation function to generate the activation gradient for the CQ system. The model gives a much better, quantitatively good fit to the GBD data in most of its target areas, and is robust in modelling a number of basic features of GBD in the face of parameter variation.

Unlike CQX, the CQM model does not represent a direct development of the CQS model, but instead takes a different route, using the 'lazy' backpropagation learning rule to generate a dynamic activation gradient. As a first model using a new architecture, its aims are closer to those of CQS than CQX - to prove the sequencing mechanism and to qualitatively capture the 'core' GBD effects. Like CQX, however, it adds the ability to model CV preservation. Simulations with the model generally show the correct qualitative behaviour.

The main differences between the three models concern the way they generate dynamic activation gradients over their letter nodes. The competitive systems which produce the final output are similar in all three cases (with the exception of the additional biasing arrangement in CQX). The differences between their qualitative and quantitative fits to the data can therefore be mainly attributed to their different activating mechanisms. The major difference in this respect between CQS and CQX is the use of an RBF activating function, which confers a greater degree of positional accuracy. The RBF activation function results in a smooth increase in support before the target position and a smooth fall-off after it, which reduces the tendency in CQS for letters to be re-activated towards the end of words, and leads to more symmetrical serial error incidence curves. Use of the same activating function with the geminate node allows geminate insertion and deletion errors to be modelled as well as movement errors. The CQM model also uses an activating system with the potential at least to provide much greater positional resolution than that of CQS, as Figure 7.5 shows. In this case, however, the improved quantitative fit achieved by CQX is not forthcoming. Some reasons why this might be the case are discussed below.

8.1.2 Shortcomings of the models

The CQS model has a number of shortcomings which are attributable to the fact that it was a first attempt to apply the CQ approach in the domain of spelling. CQM is also a 'first generation' model and likewise was not intended to provide more than a gross, qualitative fit to the GBD error pattern. Comparisons across the models may indicate which shortcomings are inherent in the CQ approach, since an attempt was made with the 'second generation' model CQX to explicitly address the shortcomings of CQS. Those which were least well addressed may indicate problems with the underlying approach rather than with details of the implementation. Additionally, CQM, while still a 'first generation' model, takes a very different approach to generating activation gradients over letters, and any problems which are common to all three models may thus indicate problems with the underlying approach.

The main area which can be identified as problematic in all three models is the incidence of deletion and insertion errors. The CQX model shows the incorrect trend in the relative incidences of deletion and substitution errors with word length, and produces very few insertions. CQM shows the correct trends with word length but produces few insertions or deletions, and both error types have serial incidence curves which are skewed towards the end of the word. CQS likewise shows skewed serial incidence curves for insertions and deletions. One of three conclusions might be drawn from these problems:

1. The use of the CQ approach for modelling the spelling output process may be inappropriate,

2. The mechanisms behind these particular error types in the models may be inappropriate, or:

3. This may be a problem which cannot be reasonably addressed by a model of the sequential output system in spelling, but may require other systems or strategic factors to be taken into account.

Taking the former possibility first, the good performance of the three CQ models in capturing the gross features of GBD errors suggests that the general approach is sound. The similarity of GBD errors to the occasional 'slips of the pen' of normal spellers reinforces this view. The second possibility has been discussed in Chapters 6 and 7, and alternative mechanisms for insertion and deletion errors have been proposed which appear promising. The third conclusion which might follow from the difficulty of modelling the correct incidence of insertion and deletion errors is that this level of detail in the error pattern of GBD patients might not be an appropriate target for a simple model of the sequential output processes in spelling. Spelling is a complex acquired skill and is not monolithic - a number of different processes and strategies are involved, some of which may be under conscious strategic control and some idiosyncratic. Glasspool, Houghton and Shallice (1995) discuss some of these strategic factors in spelling. The successful fit of the CQS, CQX and CQM models to the gross features of GBD errors indicates that modelling the sequential output processes of spelling in this manner is appropriate and informative, and the failure to obtain a perfect match for all types of error does not necessarily mean that the approach is flawed; rather, these errors may result from factors outside the remit of the models. Without a well articulated theory of the other processes involved in spelling and their interactions with the sequential output system it is better to err on the side of caution and look for ways in which these error types may be better modelled within the CQ output process. However, it is important to bear in mind that a full model of spelling must encompass these strategic and possibly idiosyncratic processes and their interaction with low-level sequencing. It may not be possible to increase the level of detail addressed by spelling models much further in the absence of theoretical progress in this direction.

8.2 Comparing the CQX and CQM models

8.2.1 Internal versus external constraints in spelling models

Both the CQX and CQM models have demonstrated the preservation of CV status in errors, with a strength comparable to the effect in GBD patients. Both the internal and external constraint approaches thus appear feasible in spelling models. A number of

issues are raised by the models concerning differences between the approaches, however.

The external approach requires a template for CV status, and there are two plausible ways in which such a template might exist within the spelling system: As a separate template or as part of the letter representation. A separate template could be a dedicated CV system with a set of templates for each possible CV structure, though this seems unlikely as the number of such structures is large and the possibilities are relatively unconstrained by comparison with speech production, where the approach has been productive (Hartley and Houghton, 1996). Another possibility is that the template is derived from phonology. This is an attractive idea with one major disadvantage - the need to correctly synchronise phonology with spelling production. This would be trivial if, for example, letters were sounded out during spelling (which might however be a more common strategy in patients with spelling difficulties than in normal spellers), but this is not generally the case. Moreover many words do not have regular correspondences between sound and letter (e.g. *yacht*). Another means by which synchronisation might be achieved is through a strategic repair process, which might operate when an error was detected or when the usual relatively unconscious spelling process failed to produce a letter. In this case a phonologically based process might be used to produce a candidate letter, which would then be in a position to feel the effect of CV status. Again, the use of such repair strategies might be more common in agraphic patients than in normal spellers, which might help explain the fact that CV preservation has not so far been noted in studies of normal speller's slips of the pen.

The strength of the external CV constraint in the CQX model can be varied freely by adjusting the level of bias from the CV template, while the level of internal constraint in the CQM model is dependent on the degree of interference or overlap between internal representations used in activating competing letters during the spelling process, which is a complex property of the model and is not easy to adjust in isolation. However, there is no reason in principle why the degree of constraint in an internal constraint system should not be as accessible as the bias level in CQX; it is the multi-layer aspect of CQM which makes this factor difficult to control, rather than the use of internal constraints per se.

8.2.2 Single layer versus multi layer CQ models

The most interesting difference between the CQX and CQM approaches is the contrast between single layer and multi-layer architectures. As both models use a similar localist output layer the difference is mainly in the letter activating system rather than the dynamics of the output system. A number of interesting issues arise in comparing the two approaches.

Output competition levels and multiple word capability

The multi-layer CQM model uses a powerful learning algorithm which has the potential to confer the same advantage as the RBF activating function in the CQX model - greater positional discrimination by comparison with CQS. (Figure 7.5 shows that the network itself is sufficiently powerful to effectively remove all competitors at each time-step if it is allowed to by the training regime). However, it is apparent when comparing Figures 6.13 and 7.4 that the dynamic behaviour of CQM is appreciably less regular than that of CQX.

One reason for this is that the CQM network handles multiple words, and there is thus a greater need for it to satisfy multiple constraints with limited resources. Single layer models would require more resources (e.g. multiple I-E node pairs) to allow the representation of multiple words, so the comparison between the two approaches is not entirely fair in this respect. The multi-layer approach allows multiple words to be very easily integrated into the model. The single-layer models reduce to a separate system for each word to be represented, so the incorporation of support for multiple words is not a particularly interesting aim when implementing such models (although interactions between known words would be possible in the single-layer versions using an interactive activation arrangement).

A second reason for the more predictable dynamic behaviour of the CQX model is the emergent nature of the activating signal in the multi-layer model. The CQM model relies on the activation profiles which emerge from the backpropagation learning regime, and it is difficult to relate the final activation profiles to the relevant training

factors, which may include not only the form of the learning rule but also the examples on which the network is trained.

Non-lexical spelling

Another major potential difference between the two approaches is the ease of integrating lexical and assembled spelling routes. The issue is a complex one, but it will be useful here to discuss one central aspect of the interaction between multiple spelling routes: The manner in which information from the assembled route is integrated into the final common output path, assumed here to be a CQ output process.

The single-layer models - CQS and CQX - assume that a separate set of I-E to item node weights exists for each sequence to be produced. The problem here is in generating an appropriate set of weights for a spelling produced on the assembled route. One possibility (proposed by Glasspool, Houghton and Shallice, 1995) is that the CQ system should be treated as a short-term memory system - a buffer in the true sense of the word - which mediates between the rapid production of a spelling by the assembled route and the slower, paced output required for writing or spelling aloud. The CQ system would be rapidly trained (using a set of temporary weights) by the assembled route, and would then reproduce the assembled spelling at the appropriate output rate. This allows the same serial output mechanism to be used for both lexical and assembled spelling and suggests a reason for the worse performance of GBD patients on non lexical material - the single presentation and possibly decaying temporary weights lead to a less robust representation for short-term stored non lexical material than for well-learned words. The CQS model demonstrates how this explanation could work by showing better overall performance with well learned words, but the general pattern of differences between words and non-words is better modelled by the use of increased noise levels for non-words, which does not correspond so well to the proposed mechanism. The single-shot learning of the CQX model makes it less obvious how a single presentation might disadvantage non-word material. However, it is possible that rapidly decaying temporary weights, possibly between the item and letter nodes, may provide a more appropriate model for non-lexical storage.

The CQM model provides a wholly different model for nonword spelling. The multi-layer framework has the potential to learn a generalised mapping from a phonemic representation to a spelling output, allowing the model to incorporate the phoneme-to-grapheme conversion system as well as the graphemic output lexicon. This should require minimal modification to the model - simply the addition of an input field representing phonology as shown in Figure 8.1. The phonological representation on this field might represent the entire word and be held constant throughout spelling, or it could be a dynamic pattern such as the shifting window of models such as NETspell (Olson and Caramazza, 1994).
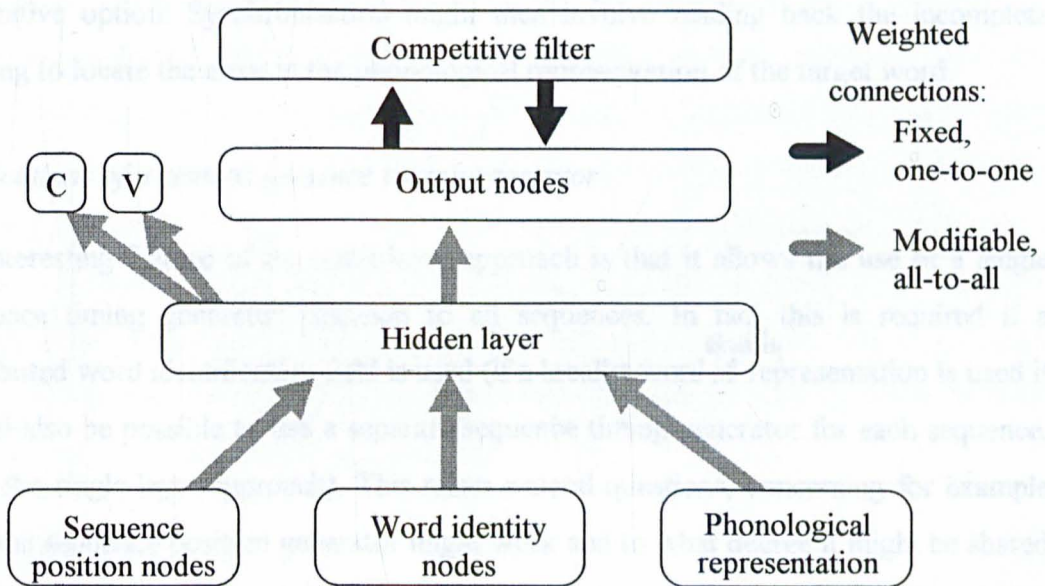
Figure 8.1. *The addition of a phonological representation of the target word at the input layer of the CQM model would give it the potential to develop an assembled phonological spelling route.*

The problem of the integration of the two routes is solved by the use of a shared hidden layer and output system by both routes. CV preservation effects would no doubt still require training the network to distinguish consonants from vowels, or the use of an external template (possibly derived from the phonological input itself). An exploration of these possibilities is beyond the scope of the current work. However, this is an area

where the differences between the two approaches are most evident and where future development will be most interesting.

A second problem with combining information from assembled and lexical routes is more general in nature, and concerns the way in which information should be combined when different candidate spellings are simultaneously available on the two routes. It is possible that repair processes during spelling production may be partly responsible for integrating input from the two routes into a single spelling. Thus when an error is detected during lexical spelling of a known word, or when the identity of a particular letter cannot be accessed, the phonological system may be used to suggest an alternative option. Synchronisation might then involve reading back the incomplete spelling to locate the error in the phonological representation of the target word.

Implications of a central sequence timing generator

An interesting feature of the multi-layer approach is that it allows the use of a single sequence timing generator common to all sequences. In fact this is required if a distributed word identification field is used (if a localist word id representation is used it would also be possible to use a separate sequence timing generator for each sequence, as in the single-layer approach). This raises several questions, concerning for example how the sequence position generator might work and to what degree it might be shared with other forms of serial behaviour. The 'Burgess and Hitch' form of sequence position representation used in the CQM model is of course only one of many possible forms, and in principle the I/E node representation used in the single layer models could be used, although this would make the mapping task rather more difficult. An interesting possibility is that the signal used to represent sequence position might be distributed over a number of systems, as for example a set of oscillators (as suggested by Church and Broadbent, 1990, and used in the OSCAR model, Brown, Preece and Hulme, submitted, see also Henson and Burgess, 1997, for a discussion of the way in which such a signal could be used for sequence timing and time-warping sequences for different learning or production rates). Such a distributed timing signal would by its nature be available for use by other sequencing systems and would be resistant to damage. A localised timing signal dedicated to the spelling system is unlikely as spelling

is a recent skill which is unlikely to have been of selective advantage for long enough for special purpose systems to have evolved to support it. It is more likely parasitic on earlier linguistic systems such as speech. The sharing of a timing signal with phonological output processes might provide a means of synchronising phonological and graphemic processes during spelling, a process which would make combining information from lexical and phonological spelling routes considerably more straightforward.

Lesioning models

Multi-layer models can make widespread use of distributed representations, and it is therefore possible to use genuine disconnection lesions on them, whereas single-layer models, which use localist representations throughout, are limited for practical purposes to noise lesions. The similar effect of noise and disconnection lesions on the CQM model is interesting. This may be partially a result of the need to average results over a number of different lesions in order because individual lesions produce very different results. A larger system, with many more connections, would enable meaningful single lesion studies to be made, which would be more interestingly different from the noise lesions available on the single-layer models.

Limitations of the backpropagation training method

The backpropagation training rule has a number of problems when compared with training rules for single-layer networks. Its biological plausibility is generally considered dubious, and it has problems with the rapid integration of new material. While a full treatment of these issues would not be appropriate here, some comments can be made on possible solutions to the practical problems.

Training a backpropagation network is notoriously slow. Because a single set of connections must come to represent multiple input to output mappings training requires a slow process of incremental weight adjustment in order to allow complex multiple constraints to be satisfied. This leads to two problems in the current context: A new word will require a long period of training before it can be incorporated into the network's vocabulary, whereas people can typically learn a new spelling rapidly, and

the introduction of new words to a network which has already acquired a vocabulary may result in catastrophic interference with the information already acquired by the network.

The problem of immediate retention of a new word on a single presentation might be addressed by the incorporation of a single-layer system into the CQM framework, which is discussed below. The occurrence of *catastrophic interference* - the loss of previously learned information when additional exemplars are added to an already learned set of input-output relations in a multi-layer network - is well known (McCloskey and Cohen, 1989). A number of possible solutions to the problem of catastrophic interference have been suggested, involving either a reduction in the overlap between the patterns stored in the system (Sharkey and Sharkey, 1995) or the use of two separate memory systems which can shift information between each other allowing gradual incorporation of new input-output mappings alongside the existing contents of the network (MClelland, McNaughton and O'Reilly, 1995, French, 1997).

Single-layer systems use a localist representation throughout, and suffer from neither of these difficulties. New words can be rapidly learned and do not interfere with previously learned words. This is a major advantage of the single-layer approach.

8.2.3 Integrating the CQX and CQM models

The CQX and CQM models have been developed and presented as alternative solutions to the same problem. It is interesting, however, to consider ways in which the different ideas developed in the two models might be applied together. Two ways in which such an approach might be useful will be considered here: Combining single and multi-layer approaches by adding a single-layer route to a CQM model, and adding an external constraint system to the CQM model.

1. Combining single and multi-layer approaches

It is possible to view single-layer and multi-layer models as operating on complimentary tasks within the same system, with the single-layer part of the system dealing with short-term retention of sequences while the multi-layer subsystem deals with long-term

storage, as shown in Figure 8.2. This corresponds to adding a set of temporary weights directly between the sequence position input field and the output layer of a CQM-like architecture. Integrated in this way, the single-layer subsystem would be limited to short-term retention because the single sequence position field would need to implicitly represent the identity of every sequence to be produced by the subsystem, as only one sequence at a time may be stored in this single layer of weights. However, if the weights are allowed to decay rapidly interference between different sequences need not be a problem, and the ability to acquire a new sequence in a single presentation would address one of the problems of the multi-layer approach. A novel sequence could be maintained by rehearsal following a single presentation while the much slower learning procedures required by the multi-layer route consolidated it into long-term memory (perhaps using one of the procedures suggested by MClelland, McNaughton and O'Reilly, 1995, or French, 1997). It does not appear that much work has been done on the short-term retention of novel spellings. However, the same general approach would apply equally to speech production. The work of Hartley and Houghton (1996) on the modelling of speech production using a CQ system with an external syllable constraint would be directly relevant to such a model, which could incorporate an external syllable template at the output layer in the same way as the CV template of CQX, as an adjunct to or replacement for the 'softer' acquired internal constraint of CQM.
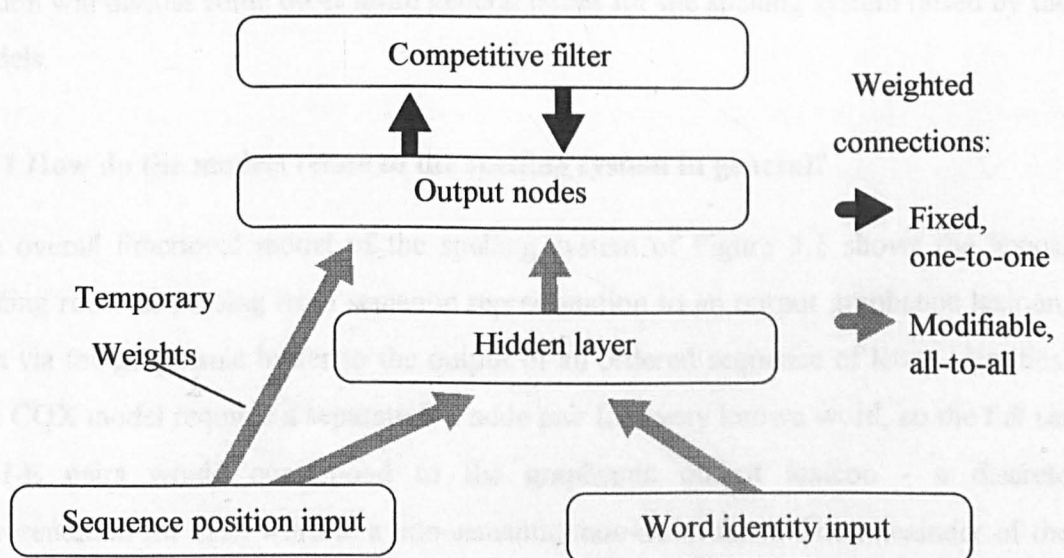

Figure 8.2. *Possible scheme to integrate single and multi-layer sequencing systems.*

2. External constraints in the CQM approach

Because the CQM model was suggested by the idea of an internal constraint system, internally mediated serial constraints have been a central part of its design. There is no reason however why the information supplied to the input layer of the model should not include an explicit serial template like that of CQX. This might correspond to moving the C and V nodes of Figure 7.1 from the output to the input layer. Given that the model was unable to acquire a strong internal CV constraint without the presence of the C and V nodes at the output the addition of an external CV template seems well motivated. As with the CQX model, it would seem reasonable to assume that such information may be available from phonological processes, although the production of this phonological information would have to be synchronised in some way with the progress of spelling production.

8.3 Implications for modelling the spelling system

So far the performance of the models in matching the disrupted spelling of GBD patients has been considered. However, an important point of the modelling effort is to gain insights into the operation of the spelling system in its undamaged state. The performance of the models has already suggested a general conclusion - that the sequential output system in spelling involves a dynamic process similar to CQ. This section will discuss some other more general issues for the spelling system raised by the models.

8.3.1 How do the models relate to the spelling system in general?

The overall functional model of the spelling system of Figure 3.1 shows the lexical spelling route as passing from semantic representation to an output graphemic lexicon, then via the graphemic buffer to the output of an ordered sequence of letter identities. The CQX model requires a separate I-E node pair for every known word, so the full set of I-E pairs would correspond to the graphemic output lexicon - a discrete representation for each word in a non-semantic, non-serial form. The remainder of the CQX model - the connections from I and E nodes to item and letter nodes and the competitive filter - would then correspond to the graphemic buffer. The mapping for

the CQM model is somewhat less clear. The model requires a unique pattern on the word id field for each known word, but this could be the representation generated by the semantic system (or perhaps only part of it). In this case it is not clear where the output lexicon starts and finishes - is the hidden layer part of it? Certainly it contains a representation pertaining to the identity of words, but it is also part of the sequence generation system and therefore presumably of the graphemic buffer. The functional blocks of Figure 3.1 may thus not map well onto the functional blocks of the CQM model.

8.3.2 A graphemic buffer?

A related issue is the question of what the CQ approach has to say about the notion of a graphemic buffer. In 'box-and-arrow' models a buffer is commonly viewed as a form of short-term storage - an device which can be loaded with some contents and then allow the contents to be read, possibly at a different rate, at a later time. While the generation of a serially ordered output from a parallel input representation is one use for buffers in such models, the notion of a buffer implies a storage capability. In CQ models (as in some other approaches to serial order generation) the notion of storing and then reading out a sequence is less appropriate. The sequential output is generated on-line by the interactions between competing and inhibited items. The term 'buffer' may thus not be appropriate for such a mechanism, which should be viewed as performing a purely a serial output generation task.

Whether the use of a buffer component in this way in box-and-arrow diagrams is functionally incorrect will depend on what additional assumptions may be made. It is probably not reasonable to assume that a sequence may be stored for any length of time in a CQ subsystem, nor that the 'buffer' may be 'loaded' in a piecemeal manner. Perhaps the most important point is that the buffer does not contain a representation of a sequence in some abstract form which can be stored or moved to another part of the system. The sequence representation does not exist independently of the sequence generator, which contains the domain-specific knowledge required to construct it.

There are two issues here - is it reasonable to retain the component currently labelled 'buffer' in a box-and-arrow diagram when it s task is being carried out by a CQ system, and if so is it reasonable to retain the label 'buffer'? CQ systems require their own representation of the competing output items, so it seems reasonable that a 'box' of some sort is required for them. However the use of the term 'buffer' is probably misleading. A term like 'sequence generator' would be more appropriate.

The use of a single-layer model with temporary weights for short-term sequence storage, on the other hand, constitutes a true graphemic buffer. However, as the CQS, CQX and CQM models show, this is not necessary for parallel to serial conversion. It may be required for short-term retention of novel spellings or for the integration of assembled phonological spellings into the general spelling output system.

8.3.3 Predictions

As well as explaining established data a good model should allow testable predictions to be made. A problem with computational models in psychology is that empirical work is generally so much better developed than current models that the available data outstrips the scope of the models, allowing few novel predictions to be made. This is to some extent true of the models advanced in this thesis - even the most developed, CQX, still leaves much data unaddressed for future work. However, the CQ framework implies some general constraints on sequence generation which constitute predictions about any mode of behaviour which uses it, as follows:

1. **Repetition.** The "select and inhibit" dynamics of CQ work against perseverative behaviour. Repeated items are thus more difficult to generate within the framework, and this is particularly the case for doubled items which require special treatment. The CQ approach therefore predicts that repeated items will facilitate errors, and more specifically that errors will tend to involve the second occurrence of an item, when its representation is likely to be refractory following the first occurrence. This effect should be stronger the fewer items intervene between first and second occurrences. Similar effects are found in verbal STM (Henson, in press, b, finds repetition inhibition for all nonadjacent repeated items, but repetition facilitation, i.e.

better recall than control items, for immediate repeats. This may be explicable by the geminate mechanism proposed in the spelling models). If a sequencing mechanism similar to CQ underlies the output processes in spelling the effect should be found in GBD spelling errors and perhaps in 'slips of the pen'.

2. **Capacity.** Under CQ, the more items which need to be encoded in a sequence, the more susceptible to disruption recall becomes, because successive items have to be stored "more closely" in the weight space (i.e., their relative positional indices, as realised by the weights, become more similar to each other). This leads to increased parallel access and competition at recall. In practice the approach is therefore limited in the length of sequence it can support in the face of disruptive noise. The prediction is therefore that longer words may have to be broken up into shorter chunks. If this occurs, the usual mechanisms responsible for the inverted-U error incidence curve should operate within chunks, resulting in increased vulnerability in the middle of each chunk.

8.4 Conclusion

The dynamic model underlying the CQ approach to serial order has already proved successful in explaining the general forms of error which subjects make on various serial tasks. The three models reported in this thesis have applied the CQ approach to spelling, and have demonstrated that the basic form of the errors made by GBD patients - the effect of word length, the basic error types, and the form of the serial error incidence curve - can be qualitatively captured by a variety of CQ-based approaches. Quantitative modelling of these effects has also been successful, with the exception of two particular types of error - insertions and deletions - which it has been argued may require a slightly different approach to that taken here. However there is nothing to suggest that even these error types cannot be explained within the general framework of CQ.

The interesting class of errors involving double letters, both in spelling and typing, in which the property of doubling dissociates from letter identity, can be both qualitatively and to some extent quantitatively captured by a simple special purpose 'geminate'

mechanism. This approach has been taken in other models but the underlying structure of the CQ approach, and especially its use of inhibition following output selection to drive sequence production forwards, provides a principled explanation for the need for such a mechanism.

The issue of serial constraints has been treated in a general way with the proposal of the internal / external distinction, and it has been suggested that both types of constraint system can be applied to CQ models additively- either by including a separate external template or by modifying the representations used internally by the model. This has been demonstrated in the case of external constraints in the CQX model, which operates as an unconstrained model with the CV bias parameter set low, and with CV constraints with a high bias setting, and in the case of internal constraints in the CQM model, which exhibits internal constraints when it is taught to distinguish consonants from vowels.

The CQM model demonstrates the feasibility of generalising the CQ approach to sequencing in multi-layer network models. The move to a MLP architecture has introduced difficulties, particularly with understanding and controlling the shape of item activation profiles during sequence generation. However, the amount of existing work in the paradigm and the possibilities for combining multiple input fields suggest that the approach is worth developing.

Competitive Queuing thus seems an appropriate framework for modelling the output stages of the spelling system. Approaching the regularities in errors on serial tasks from the point of view of constraints acting on a CQ sequencing system shows every prospect of being a productive framework for understanding the operation of low-level serial systems.

References

Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology* **76**, 129-140.

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). Human memory: A proposed system and its control processes. In Spence, K.W. (Ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol 2 (pp. 90-191). New York: Academic Press.

Baars, B. J. & Motley, M. T. (1976). Spoonerisms as sequencer conflicts: Evidence from artificially elicited errors. American Journal of Psychology, **89**, 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., & Hitch, G. L. (1974). Working Memory. In Bower, G.H (Ed.) *Recent Advances in the Psychology of Learning and Motivation*, Vol. VIII (pp. 47-90). New York: Academic Press.

Baddeley, A. D. & Hitch, G. L. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and Performance*, Vol. VI, 647-667. Hillsdale, N.J.: Lawrence Erlbaum Associates Ltd.

Baddeley, A.D, Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour* **14**, 575-589.

Ballard, D. H. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, **9**, 67-120.

Beauvois, M., F., & Derouesne, J. (1981). Lexical or orthographic agraphia. *Brain*, **104**, 21-49

Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the mind.* Cambridge, Mass: Basil Blackwell.

Benguerel, A. P. & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica,* **30,** 41-55.

Besner, D. & Davelaar, E. (1982). Basic processes in reading: two phonological codes. *Canadian Journal of Psychology,* **36,** 701-711.

Bjork , E. L. & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behaviour* **13,** 80-97.

Bjork, R. A. & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology* **6,** 173-189.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin* **109,** 204-223.

Brown, G. D. A & Loosemore, R. P. W. (1994). Computational approaches to normal and impaired spelling. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of Spelling: Theory, Process and Intervention* (pp. 319-336). Chichester: John Wiley and Sons.

Brown, G. D. A., Preece, T. & Hulme, C. (Submitted). *Oscillator-based memory for serial order.*

Brown, G. D. A. & Vousden, J. I. (In press). Adaptive analysis of sequential behaviour: Oscillators as rational mechanisms. In M. Oaksford and N. Chater (Eds.) *Rational Models of Cognition.* Oxford: Oxford University Press.

Brown, R. & McNeill, D. (1966). The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behaviour,* **5,** 325-337.

Burgess, N., & J. G. Hitch (1992). Towards a network model of the articulatory loop. *Journal of Memory and Language,* **31,** 429-460.

Burgess, N. & Hitch, G. J. (1996). A connectionist model of STM for serial order. In S. E. Gathercole (Ed.) *Models of Short-Term Memory* (pp. 51-72). Hove: Psychology Press.

Burwitz, L. (1974). Proactive interference and directed forgetting in motor short-term memory. *Journal of Experimental Psychology,* **102**, 799-805.

Caramazza, A. & Miceli, G. (1990). The structure of graphemic representations. *Cognition,* **37**, 243-297.

Caramazza, A., Miceli, G., Villa, G. & Romani, C. (1987). The role of the graphemic buffer in spelling: evidence from a case of acquired dysgraphia. *Cognition,* **26**, 59-85.

Christiansen, M. H. (1997). Improving learning and generalisation in neural networks through the acquisition of multiple related functions. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist Representations.* Proceedings of the 4th Neural Computation and Psychology Workshop (pp. 58-70). London, Springer-Verlag..

Church, R. M. & Broadbent, H. (1990). Alternative representations of time, number and rate. *Cognition,* **37**, 55-81.

Conrad R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology,* **55**, 75-84.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behaviour* **4**, 161-169.

Conrad, R. & Hull, A. J. (1964), Information, acoustic confusion and memory span. *British Journal of Psychology,* **55**, 429-432.

Cooper, R. & Shallice, T. (1997): Modelling the selection of routine action: Exploring the criticality of parameter values. In Shafto, M. G., & Langley, P. (Eds.), Proceedings of the 19th Annual Conference of the Cognitive Science Society (pp.131-136). Stanford, CA.

Cooper, R. & Shallice, T. (In submission). Contention Scheduling and the control of routine activities.

Cooper, R., Shallice, T., & Farringdon, J. (1995). Symbolic and continuous processes in the automatic selection of actions. In J. Hallam (Ed) *Hybrid Problems, Hybrid Solutions, Frontiers in Artificial Intelligence and Applications* (pp. 27-37). Amsterdam: IOS Press.

Crowder, R. G. (1972). Visual and auditory memory. In J. Kavanagh & I. Mattingley (Eds.) *Language by Ear and by Eye* (pp. 251-275). Cambridge, Mass.: The MIT Press.

Cubelli, R. (1991). A selective deficit for writing vowels in acquired dysgraphia. *Nature*, **353**, 258-260.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 222-233.

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

Dell, G. S. (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. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, **104**, 123-147.

Ebenholtz, S. M. (1963). Position mediated transfer between serial learning and a spatial discrimination task. *Journal of Experimental Psychology*, **65**, 603-608.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review* **89**, 627-661.

Ellis, A. W. (1979). Slips of the pen. *Visible Language*, **13**, 265-282.

Ellis, A. W. (1982). Spelling and writing (and reading and speaking). In A. W. Ellis (Ed.), *Normality and pathology in cognitive function* (pp. 251-275). London: Academic Press.

Ellis, N. .C. & Henneley, R. .A. (1980), A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, **71**, 43-51.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, **14**, 179-211.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.) *Coding Processes in Human Memory* (pp. 161-190). Washington DC: V. H. Winston & sons.

Fay, D. & Cutler, A., (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, **8**, 505-520

French, R. M. (1997). Selective memory loss in aphasics: An insight from pseudo-recurrent connectionist networks. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist Representations*. Proceedings of the 4th Neural Computation and Psychology Workshop (pp183-195). London, Springer-Verlag.

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

Fromkin, V. A. (Ed.) (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Fromkin, V. A. (Ed.) (1980). *Errors in Linguistic Performance*. New York: Academic Press.

Fuchs, A. H. (1969). Recall for order and content of serial word lists in short-term memory. *Journal of Experimental Psychology*, **82**, 14-21.

Glanzer, M. (1972). Storage mechanisms in recall. In Bower, G.H. (Ed.) The Psychology of Learning and Motivation: *Advances in Research and Theory,* Vol. V (pp. 129-193). New York: Academic Press.

Glanzer, M. & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behaviour,* **5,** 351-360.

Glasspool, D.W. (1995). Competitive queueing and the articulatory loop: An extended network model. In J. Levy, D. Bairaktaris, J. Bullinaria, & P. Cairns (Eds.), *Connectionist Models of Memory and Language* (pp. 5-30). London: UCL Press.

Glasspool, D. W. & Houghton, G. (1997). Dynamic representation of structural constraints in models of serial behaviour. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist Representations.* Proceedings of the 4th Neural Computation and Psychology Workshop (pp. 269-282). London, Springer-Verlag.

Glasspool, D. W., Houghton, G., & Shallice, T. (1995). Interactions between knowledge sources in a dual-route connectionist model of spelling. In L. S. Smith and P. J. B. Hancock (Eds.) *Neural Computation and Psychology* (pp. 209-226). London: Springer-Verlag.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin,* **99,** 221-228.

Grossberg, S. (1978). A theory of human memory: Self-organisation and performance of sensory-motor codes, maps and plans. *Progress in theoretical biology,* **5,** 233-302.

Grudin, J. T. (1983). Error patterns in skilled and novice transcription typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 95-120). New York: Springer-Verlag.

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.

Hartley, T. & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language* **35,** 1-31.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, **13**, 644-655.

Healy, A. F. (1975). Coding of spatial-temporal patterns in short-term memory. *Journal of Verbal Learning and Verbal Behaviour* **14**, 481-495.

Hebb, D. O. (1949). *The Organisation of Behaviour*. New York: Wiley

Henson, R. N. A (In press, a.) Short-term memory for serial order: The Start-End model. *Cognitive Psychology*.

Henson, R. N. A. (In press, b.) Item repetition in short-term memory: Ranschburg revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Henson, R. N. A. (Submitted). Positional information in short-term memory: Relative or absolute?

Henson, R. N. A & Burgess, N. (1997). Representations of serial order. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist Representations*. Proceedings of the 4th Neural Computation and Psychology Workshop (pp. 283-300). London, Springer-Verlag.

Henson R. N. A, Burgess, N, & Frith, C. (In preparation). Phonological short-term memory: an fMRI study.

Henson, R. N. A., Norris, D. G., Page, M. P. A. & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, **49A**, 80-115.

Hillis, A. & Caramazza, A. (1989). The graphemic buffer and attentional mechanisms. *Brain & Language*. **36**, 208-235.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence* **40**, 185-234.

Hinton, G. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. Psychological Review **98**, 74-95. *Deep dyslexia model lesion studies.*

Hitch, G. J. (1974). Short-term memory for spatial and temporal information. *Quarterly Journal of Experimental Psychology* **26**, 503-513.

Hitch, G. J., Burgess, N., Towse, J. N. & Culpin, V. (1996). Temporal grouping effects . in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*, **49**, 116-139.

Hotopf, W. H. N. (1980). Slips of the pen. In U. Frith (Ed.), *Cognitive processes in spelling.* London: Academic Press. 287-307.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, M. & Zock (Eds*.), Current Research in Natural Language Generation* (pp. 287-319). London: Academic Press.

Houghton, G., (1994a), Some formal variations on the theme of competitive queueing. Internal Technical Report, *UCL-PSY-CQ1*. Dept. of Psychology, University College London.

Houghton, G. (1994b). Inhibitory control of neurodynamics: Opponent mechanisms in sequencing and selective action. In M. Oaksford and G. D. A. Brown (Eds.) *Neurodynamics and Psychology.* London: Academic Press.

Houghton, G., Glasspool, D. & Shallice, T. (1994). Spelling and serial recall: Insights from a competitive queueing model. In G.D.A. Brown & N.C. Ellis (Eds*.), Handbook of Spelling: Theory, Process and Intervention* (pp. 365-404). Chichester: John Wiley and Sons.

Houghton, G. & Hartley, T. (1996). Parallel models of serial behaviour: Lashley revisited. *Psyche.* psyche-95-2-25-lashley-1-houghton.

Hulme, C., Maughan, S. & Brown, C.D.A. (1991). Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to short-term span. *Journal of Memory and Language*, **30**, 685-701.

Jahnke, J. C. (1969). The Ranschburg effect. *Psychological Review*, **76**, 592-605.

Johnson, G. J. (1991). A distinctiveness model of serial learning. *Psychological Review*, **98**, 204-217.

Jonsdottir, M., Shallice, T., & Wise, R. (1996). Language-specific differences in graphemic buffer disorder. *Cognition*, **59** 169-197.

Jordan, M., (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (pp. 10-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Katz, R. (1991). Limited retention of information in the graphemic buffer. *Cortex*, **27**, 111-119.

Kay, J. & Hanley, R. (1994). Peripheral disorders of spelling: The role of the graphemic buffer. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of Spelling: Theory, Process and Intervention* (pp. 295-315). Chichester: John Wiley and Sons.

Koriat, A. & Lieblich, I. (1974). What does a person in a TOT state know that a person in a don't know state doesn't know. *Memory and Cognition*, **2**, 647-655.

Koriat, A. & Lieblich, I. (1975). Examination of the letter serial position effect in the "TOT" and the "don't know" states. *Bulletin of the Psychonomic Society*, **6**, 539-541.

Kruschke, J. K. (1992). ALCOVE: A connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Kucera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

Lashley, K. S. (1951). The problem of serial order in behaviour. In F. Beach, D. Hebb, C. Morgan and H. Nissen (Eds.) (1960) *The Neuropsychology of Lashley* (pp. 506-528). New York: McGraw-Hill.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behaviour*, **16**, 395-418.

Lee, C. L. & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 149-169.

Lessenberry, D. D. (1928). *Analysis of errors*. Syracuse, NY: Smith and Corona Typewriters, School Dept. Reprinted in A. Dvorak, N. Merrick, W. Dealey and G. Ford (1936) *Typewriting Behaviour*. New York: American Book Company.

Levelt, W. (1989). *Speaking*. Cambridge, MA. MIT Press.

Lewandowsky, S. & Murdock, B. B. (1989). Memory for serial order. *Psychological Review* **96**, 25-57.

Link, K. & Caramazza, A. (1994). Orthographic structure and the spelling process: A comparison of different codes. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of Spelling: Theory, Process and Intervention* (pp. 261-294). Chichester: John Wiley and Sons.

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 organization of perception and action*. New York: Springer Verlag.

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986). *Parallel distributed processing. Explorations in the microstructure of cognition*. Cambridge Mass.: M.I.T. Press.

Magill, R. A. & Dowell, M. N. (1977). Serial position effects in motor short-term memory. Journal of Motor Behaviour, **9**, 319-323.

Martin, N., Dell, G. S., Saffran, E. M. & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia - testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, **47**, 4, 609-660

McCloskey, M. & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning and Motivation*, **24**, 109-165.

McCloskey, M., Badecker, W., Goodman-Shulman, R. & Aliminosa, D. (1994). The structure of graphemic representations in spelling: Evidence from a case of acquired dysgraphia. *Cognitive Neuropsychology*. **11**, 341.

MClelland, J., McNaughton, B. & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**. 419-457.

Miceli, G., Silveri, M. C. & Caramazza, A. (1985). Cognitive analysis of a case of pure dysgraphia. *Brain and Language*, **25**, 187-196.

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.

Milner, P. M., (1961). A neural mechanism for the immediate recall of sequences. *Kybernetic_1*, 76-81.

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

Moody, J. & Darken, C. (1988). Learning with localised receptive fields. In D. Touretzky, G. Hinton & T. Sejnowski (Eds.) *Proceedings of the 1988 Connectionist Models Summer School (Pittsburgh 1988)*, (pp. 133-143). San Mateo: Morgan Kaufmann.

Morton, J. (1980). The logogen model and orthographic structure. In Frith, U. (Ed.), *Cognitive approaches in spelling* (pp. 117-134). London: Academic Press.

Motley, M. T. (1973). An analysis of spoonerisms as psycholinguistic phenomena. *Speech Monographs* **40**, 66-71.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review* **67**, 16-31.

Murdock, B. B. (1962) The serial position effect of free recall. *Journal of Experimental Psychology* **64**, 482-488.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review,* **89**, 609-626.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review* **90**, 316-338.

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory and Cognition,* **23**, 631-645.

Murdock, B. B. & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology,* **74**, 137-143.

Murray, D. J. (1966). Vocalization-at-presentation and immediate recall with varying recall methods. *Quarterly Journal of Experimental Psychology,* **18**, 9-18.

Nigrin, A. L. (1993). *Neural networks for pattern recognition.* Cambridge, Massachusetts: MIT Press.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behaviour.* Center for Human Information Processing (Technical Report No. 99). University of California, San Diego.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. Reprinted in revised form in R. J. Davidson, G. E. Schwartz, &

D. Shapiro (Eds.) 1986. *Consciousness and self-regulation*, Vol. 4 (pp. 1-18). New York: Plenum Press.

Olson, A. & Caramazza, A. (1994). Representation and connectionist models: the NETspell experience. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of Spelling: Theory, Process and Intervention* (pp. 337-363). Chichester: John Wiley and Sons.

Page, M. P. A. (1993). *Modelling aspects of music perception using self-organising neural networks*. Unpublished doctoral dissertation, University of Wales.

Page, M. P. A. & Norris, D. G. (1997). A localist implementation of the primacy model of immediate serial recall. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist Representations*. Proceedings of the 4th Neural Computation and Psychology Workshop (pp. 316-330). London, Springer-Verlag.

Page, M. P. A. & Norris, D. G. (submitted). The primacy model: A new model of immediate serial recall.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. Cognitive Neuropsychology, **10**, 377-500.

Posteraro, L., Zinelli, P. & Mazzucchi, A. (1988). Selective impairment of the graphemic buffer in acquired dysgraphia: a case study. *Brain and Language*, **35**, 274-286.

Postman, L. & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, **17**, 132-138.

Quinlan, P. T. (1993). *The Oxford Psycholinguistic Database*. Oxford: Oxford University Press.

Rabbitt, P. M. A. (1978). Detection of errors by skilled typists. *Ergonomics*, **21**, 945-958.

Rosenbaum, D. A. (1991). *Human Motor Control*. San Diego: Academic Press.

Rosenbaum, D. A, Kenny, S. B. & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 86-102.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by back-propagating errors. *Nature* **323**, 533-536.

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

Sevald, C., Dell, G. & Cole, J. (1995). Syllable structure in speech production - are syllables chunks or schemas? *Journal of Memory and Language* **34**, 807-820.

Seymour, P. H. K. & Porpodas, C. D. (1980). Lexical and non-lexical processing of spelling in dyslexia, In U. Frith (Ed.), *Cognitive Processes in Spelling* (pp. 443-474). London: Academic Press.

Shaffer, L. H. (1975). Control processes in typing. *Quarterly Journal of Experimental Psychology*, **27**, 419-432.

Shaffer, L. H. (1976). Intention and performance. *Psychological Review*, **83**. 375-393.

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain*, **104**, 413-429.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shallice, T., Glasspool, D. W. & Houghton, G. (1995). Can neuropsychological evidence inform connectionist modelling? Analyses of spelling. *Language and Cognitive Processes* **10**, 195-225.

Sharkey, N. & Sharkey, A. (1995). An analysis of catastrophic interference. *Connection Science*, **7**. 301-329.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper and E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garret* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H. A. (1974). How big is a chunk? *Science,* **183,** 482-488.

Slamecka, N. J. (1964). An inquiry into the doctrine of remote associations. *Psychological Review,* **71,** 61-76.

Stemberger, J. P. (1985), An interactive activation model of language production. In W.W. Ellis (Ed.), *Progress in the psychology of language,* Vol. 1 (pp. 143.186). Hillsdale, NJ: Erlbaum.

Sternberg, S. (1966). High-speed scanning in human memory. *Science,* **153,** 652-654.

Sternberg, S., Monsell, S., Knoll, R. L. & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117-152). New York: Academic Press.

Tainturier, M. J., & Caramazza, A. (1996). The status of double letters in graphemic representations. *Journal of Memory and Language,* **35,** 53-73.

Tweney, R. D, Tkacz, S. & Zaruba, S. (1975). Slips of the tongue and lexical storage. *Language and Speech,* **18,** 388-396.

Tzeng, O. J. L. (1973). Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behaviour,* **12,** 436-439.

Vousden, J. I. (1996). *Serial control of phonology in speech production.* Unpublished doctoral dissertation, University of Warwick, England.

Vousden, J. & Brown, G. D. A. (1997). To repeat or not to repeat: The time course of response. In J. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Connectionist*

Representations. Proceedings of the 4th Neural Computation and Psychology Workshop (pp. 301-315). London, Springer-Verlag.

Watkins, M. .J. (1977). The intricacy of memory span. *Memory and Cognition*, **5**, 529-534.

Watkins, M. J. & Peynircioglu, Z. F. (1983). Three recency effects at the same time. Journal of Verbal Learning and Verbal Behaviour, **22**, 375-384.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Quarterly Journal of Experimental Psychology*, **19**, 97-102.

Wilberg R. B. (1990). The retention and free recall of multiple movements. *Human Movement Science*, **9**, 437-479.

Wing, A. M., & Baddeley, A. D. (1980). Spelling errors in handwriting: A corpus and a distributional analysis. In Frith, U. (Ed.), *Cognitive processes in spelling* (pp. 251-285). London: Academic Press.

Wrisberg, J. (1975). The serial position effect in short-term motor retention. *Journal of Motor Behaviour*, **7**, 289-295.

Young, R. K. (1961). The stimulus in serial verbal learning. American Journal of Psychology, **74**, 517-528.

Young, R. K. (1962). Tests of three hypotheses about the effective stimulus in serial learning. Journal of Experimental Psychology, **63**, 307-313.

Young, R. K. (1968). Serial Learning. In T. Dixon and D. Horton (Eds.) *Verbal Behaviour and General Behaviour Theory* (pp. 122-148). Englewood Cliffs, NJ: Prentice Hall.

Yu, A. C. & Margoliash, D. (1996). Temporal Hierarchical Control of Singing in Birds. *Science* **273**, 1871-1875.