

Vocabulary Acquisition and Verbal Short-Term Memory: Computational and Neural Bases

Prahlad Gupta and Brian MacWhinney

Carnegie Mellon University

In this paper, we explore the hypothesis that human vocabulary acquisition processes and verbal short-term memory abilities utilize a common cognitive and neural system. We begin by reviewing behavioral evidence for a shared set of processes. Next, we examine what the computational bases of such a shared system might be and how vocabulary acquisition and verbal short-term memory might be related in mechanistic terms. We examine existing computational models of vocabulary acquisition and of verbal short-term memory, concluding that they fail to adequately relate these two domains. We then propose an alternative model which accounts not only for the relationship between word learning and verbal short-term memory, but also for a wide range of phenomena in verbal short-term memory. Furthermore, this new account provides a clear statement of the relationship between the proposed system and mechanisms of language processing more generally. We then consider possible neural substrates for this cognitive system. We begin by reviewing what is known of the neural substrates of speech processing and outline a conceptual framework within which a variety of seemingly contradictory neurophysiological and neuropsychological findings can be accommodated. The linkage of the shared system for vocabulary acquisition and verbal short-term memory to neural areas specifically involved in speech processing lends further support to our functional-level identification of the mechanisms of vocabulary acquisition and verbal short-term memory with those of language processing. The present work thus relates vocabulary acquisition and verbal short-term memory to each other and to speech processing, at a cognitive, computational, and neural level. © 1997 Academic Press

Learning new words (i.e., vocabulary acquisition) is one of the most crucial processes in human development. Without a system for learning words we could never acquire language, and without language, human culture could not be developed and could not be maintained. A second critical and charac-

We thank Tom Hartley for providing program code for the computer model described in Hartley and Houghton (1996). We also thank Dave Plaut and three anonymous reviewers, for helpful comments on earlier versions of this paper.

Address correspondence and reprint requests to Prahlad Gupta, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

teristic human faculty is the ability to retain sequences of words in short-term memory. Without this additional mnemonic ability, it would be impossible to understand anything but the simplest of sentences. Difficulties in learning new words or in remembering sequences of words can place human infants at serious risk for abnormal language development, and deficits in these foundational abilities can lead to cumulating backlogs in language development.

Unfortunately, experimental psychology has paid little attention to the investigation of processes involved in learning new words, and virtually nothing is known about the abilities and component processes underlying this fundamental ability. Research in the area of child language development has served primarily to document either the role of constraints on word learning or the outcome of word learning under various conditions rather than to elucidate the cognitive processes involved (e.g., Au & Glusman, 1990; Baldwin & Markman, 1989; Callanan & Markman, 1982; Carey, 1978; Carey & Bartlett, 1978; Dickinson, 1984; Dollaghan, 1985, 1987; Holdgrafer & Sorensen, 1984; Keefe, Feldman, & Holland, 1989; Leonard, Schwartz, Morris, & Chapman, 1981; Markman, 1984, 1989; Markman & Hutchinson, 1984; Merriman & Bowman, 1989; Merriman & Schuster, 1991; Mervis, 1984; Mervis & Pani, 1980; Mervis, Golinkoff, & Bertrand, 1994; Nelson & Bonvillian, 1973; Oviatt, 1980, 1982; Schwartz & Leonard, 1984).

Recent evidence suggests, however, that human vocabulary acquisition processes and aspects of human working memory may be closely related (Gathercole & Baddeley, 1989, 1990, 1993; Papagno, Valentine, & Baddeley, 1991). This work has been conducted within the framework of the working memory model (Baddeley, 1986), which therefore offers a valuable theoretical perspective for investigation of vocabulary acquisition. What makes it an especially useful framework is the further fact that a good deal is known about the neural substrates of verbal short-term memory, both from neuropsychological investigation (Baddeley, Papagno, & Vallar, 1988; Shallice & Vallar, 1990; Waters, Rochon, & Caplan, 1992) and from neuroimaging studies (Blamire et al., 1993; Cohen et al., 1993; Grasby et al., 1993; Jonides et al., 1993; Paulesu et al., 1993a; Paulesu, Frith, & Frackowiak, 1993b; Petrides, Alivisatos, Meyer, & Evans, 1993). This together with increasing neuroimaging investigation of the substrates of phonological processing (Petersen, Fox, Posner, Mintun, & Raichle, 1989; Demonet et al., 1992; Howard et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992), makes it possible to identify specific neural circuitry that may be involved in vocabulary acquisition processes and working memory.

The search for shared mechanisms underlying verbal short-term memory and vocabulary acquisition is important for several reasons. First, it offers a new processing-oriented approach to examining vocabulary acquisition. Second, exploration of word learning can illuminate the relations between short- and long-term memory systems. Third, examination of this linkage of two critical human facilities would provide us with a richer way of under-

standing the details of language learning. Fourth, to the extent that the neural substrates of such processing can be specified, this relationship has considerable importance in investigation of brain and language. Fifth, identification of shared component processes and neural substrates in vocabulary acquisition and verbal short-term memory can illuminate the study of language disorders. Better understanding of vocabulary acquisition and verbal short-term memory both at a process level and in terms of their neural substrates is potentially important for the diagnosis and treatment of a number of patient populations, such as children with perinatal brain lesions, children with specific language impairment (SLI), and adult aphasics.

Two fundamental questions can thus be posed. First, how are vocabulary acquisition and verbal short-term memory related in terms of cognitive processing? Second, what are the brain-behavior relations underlying these two cognitive domains? This paper is intended as a step toward developing a theoretical framework relating vocabulary acquisition and verbal short-term memory in computational as well as neural terms.

The first section reviews behavioral evidence for the existence of a common processing system underlying vocabulary acquisition and verbal short-term memory, highlighting the need for a unified theoretical/computational framework within which to think about the issues. The second section lays out desiderata for an integrated account, in terms of the phenomena, both in word learning and in verbal short-term memory, that must be accounted for by an adequate model. The third and fourth sections review existing computational models of vocabulary acquisition and of working memory, respectively, showing that existing accounts provide only partial coverage of one or the other domain. The fifth section sketches the framework of a new model which addresses the relationship between vocabulary acquisition and working memory as well as a wide range of phenomena in verbal working memory. The model as described here is a proposal and has not yet been fully implemented, although implementation is in progress (Gupta, 1996). Nevertheless, because it draws on the existing models, the computational feasibility of its proposed computational mechanisms is not in doubt. As we shall see, it provides a theoretical framework within which the relationship between vocabulary acquisition and working memory can be explicated.

The new model holds that the processes of immediate serial recall are firmly embedded within the overall human speech processing system. To further ground this functional account, we turn in the sixth section to a consideration of the potential neural substrates of such a processing system. We begin by reviewing what is known of the neural substrates of speech processing and suggest a conceptual framework within which a variety of seemingly contradictory findings can be accommodated. In this conceptualization, the speech processing system is viewed as interactive in nature; input and output phonetics map respectively onto auditory cortex and Broca's area and

are more loosely coupled than input and output phonology, which map onto an interconnected system composed of other anterior and posterior areas of left perisylvian cortex; we discuss how this conceptualization differs in important ways from the classical view of anterior and posterior language areas. Finally, we map our model onto these neural substrates, thus concluding our linkage of word learning and verbal short-term memory at a behavioral, computational, and neural level.

VOCABULARY ACQUISITION AND VERBAL SHORT-TERM MEMORY: A SHARED SYSTEM?

The working memory model has been one of the most widely influential theories of short-term memory over the past 20 years (Baddeley & Hitch, 1974; Baddeley, 1986). The model has three major components: a visuo-spatial short-term memory, a verbal short-term memory, and a central executive, which controls the flow of information to and from the other components. The verbal short-term memory system has been termed the "articulatory loop" and consists of two parts (Baddeley, 1986). One subcomponent of this processing system consists of mechanisms that enable rehearsal (e.g., repeating a phone number to oneself until it can be dialed); the second subcomponent consists of a system for phonological representation of individual spoken words and their temporary storage (Baddeley, 1990b). The rehearsal process "refreshes" the decaying traces in the memory store; it is believed to rely on articulatory mechanisms and has therefore been called articulatory rehearsal (Baddeley, 1986). The memory buffer stores verbal material, but its memory traces decay within 1–2 sec, which is why a refresh mechanism is needed.

Behavioral Evidence for a Relationship

One easy way of measuring verbal short-term memory is to ask subjects to recall lists of words. In the immediate serial recall (ISR) task, the subject is presented with sequences of unrelated verbal items (such as digits or words) and is required to recall the sequence in correct order, immediately following its presentation. Presentation of the list may be either auditory or visual. The subject may be required to respond in speech, in writing, or in some other fashion. The subject's digit span is measured as the length of the longest list that the subject can recall at some criterion of performance (e.g., correct recall on 60% of trials at that list length). This task has played a central role in development of the working memory model.

Digit span as measured by ISR is widely accepted as the standard measure of verbal short-term memory. Gathercole and Baddeley (1989) have described a group of language-disordered (specific language impairment) children, whose digit span was found to be highly correlated with their (poor) nonword repetition performance. This finding suggests that common abilities

are involved in immediate serial recall and nonword repetition and that nonword repetition can be used as an alternative to ISR to gauge verbal short-term memory abilities. To further examine their verbal short-term memory abilities, the SLI group was compared with a control population of normal children matched in terms of vocabulary and reading skills. The SLI children's performance was significantly impaired, relative to that of the controls, on nonword repetition as well as digit span. Thus, the SLI group, whose mean age was 8 years, had nonword repetition ability equivalent to that of 4-year-olds, while their vocabulary abilities were equivalent to those of 6-year-olds (the control group). These findings are important because they suggest that the same abilities are involved in immediate serial recall, nonword repetition, and vocabulary acquisition, as the SLI group's performance was impaired in all three of these areas.

Gathercole and Baddeley (1990) have also described a longitudinal study of 4- and 5-year-old normal children, in which nonword repetition scores (which index verbal working memory; Gathercole, Willis, Baddeley, & Emslie, 1994) were highly correlated (a) with receptive vocabulary scores and (b) with the time taken to learn unfamiliar names in a simulated vocabulary acquisition task. Once again, this suggests that the mechanisms underlying verbal short-term memory play an important role in vocabulary acquisition.

In a study of 9-year-old Finnish children learning English as a foreign language, Service (1992) found a close association between nonword repetition ability and English grades 2 years later. Repetition scores were not, however, correlated with arithmetic scores. This suggests that the correlation between verbal short-term memory (as gauged by nonword repetition) and foreign-language learning reflects reliance on some common processing component and not merely common reliance on general intellectual abilities.

The studies by Gathercole, Baddeley, and Service have focused on the long-term correlational effects of verbal short-term memory abilities with vocabulary acquisition. Another approach to the study of the linkage between these skills looks at processing by adults in a controlled experimental context. Papagno et al. (1991) have examined normal adult subjects' paired-associate learning, i.e., their ability to learn associations between pairs of words so as to be able to recall one member of the pair when prompted with the other. The experiment tested learning of English-English paired associates as well as learning of English-Finnish paired-associates. That is, the subjects' task was to learn to associate known English words with an unrelated phonological form, which was either (a) an English word or (b) a Finnish word. Subjects were asked to perform these tasks concurrently with one of two other tasks. In a concurrent articulation condition, subjects had to repeatedly utter the word *bla* while simultaneously performing the primary paired-associate tasks. In a concurrent finger tapping condition, subjects had to tap their finger repeatedly while simultaneously performing the primary

tasks. That is, subjects had to perform paired-associate learning (of either English–English pairings or English–Finnish pairings), while at the same time performing a concurrent interference task (either articulation or finger-tapping). Extrapolating from previous studies using concurrent articulation interference tasks (Baddeley, Thomson, & Buchanan, 1975; Salame & Baddeley, 1982; Baddeley, Lewis, & Vallar, 1984; Hanley & Broadbent, 1987; Baddeley, Papagno, & Norris, 1991), Papagno et al. (1991) predicted that vocabulary learning (learning English–Finnish pairings) would be affected more by concurrent articulation than by concurrent finger-tapping. This prediction was confirmed. There was a significant difference between the impact of the two interference tasks (concurrent articulation and finger tapping) on English–Finnish learning, but no significant difference in the impact of the two tasks on English–English learning. These findings are important because they provide further evidence that articulatory rehearsal plays a role in vocabulary acquisition and that concurrent articulation interferes with this rehearsal.

A further line of evidence comes from neuropsychological data. The neuropsychological syndrome of a ‘pure STM’ deficit involves reduced auditory–verbal short-term memory (STM) in the absence of other major language and cognitive deficits (Shallice & Vallar, 1990). Baddeley et al. (1988) have described a patient, P.V., who has a pure STM deficit. P.V. was able to learn meaningful paired associates in a familiar language. However, she was unable to learn to associate an unfamiliar word (in an unfamiliar language) with a familiar word in a familiar language, which is akin to learning a new vocabulary item. Learning in the second (vocabulary acquisition) condition involved mapping an arbitrary, novel phonological form to a known semantics. The first (paired associates) condition involved mapping of an arbitrary but known phonological form to a known semantics. Thus both conditions involved an associative mapping of a phonological form to a known semantics, but only the second condition required learning a new phonological form. The differences in P.V.’s performance between the two conditions could therefore be attributed to this difference. The fact that P.V. was a pure STM patient suggests, once again, that immediate serial recall and vocabulary acquisition may involve common processing mechanisms.

The Need for an Integrated Computational Model

To summarize, there is now considerable evidence suggesting that aspects of the rehearsal and phonological representation/storage system underlying verbal short-term memory also underlie vocabulary acquisition. The articulatory loop model suggests that there is a partnership of (A) a phonological representation/storage system and (B) speech output planning mechanisms, such that the latter can serve to generate and/or refresh representations over the former. The partnership of these two components provides a very simple

model of the system underlying performance in both verbal short-term memory and vocabulary acquisition. In immediate serial recall, this system sets up a loop of activations (rehearsal) which can maintain phonological representations of the recall stimuli in an active state. In vocabulary acquisition, likewise, a similar rehearsal loop aids the formation of a phonological representation for a new vocabulary item, by allowing the learner repeated access to the form so as to consolidate the new memory.

To clarify the nature of this system, it is important to have a computational framework for thinking about the issues. The working memory model has remained vague about the mechanisms of verbal short-term memory and about how cognitive processing in serial recall tasks might relate to any other cognitive apparatus (Monsell, 1987), a limitation acknowledged by the architects of the model who also note that this has hindered further development of the model (Baddeley & Hitch, 1994).

In view of the significance of the relationship between vocabulary acquisition and verbal short-term memory, it is surprising that little is currently known about it beyond the correlational data. It is also surprising that there have been no attempts to integrate these two fields within a unified theoretical framework. The available computational models of word learning (Grossberg, 1978; Houghton, 1990; Miikkulainen, 1990) do not directly address the phenomena of verbal short-term memory. Further, the available models of verbal short-term memory (Bairaktaris & Stenning, 1992; D. Bairaktaris & K. Stenning, manuscript in preparation; Brown, 1989; Brown & Hulme, 1995; Burgess & Hitch, 1992; Hartley & Houghton, in press; Houghton, 1993; Houghton, in press; D. W. Glasspool, manuscript in preparation; H. B. Richman & H. A. Simon, manuscript in preparation) do not consider the process of word learning (but see Burgess, 1995).

A framework that would specify computational mechanisms linking verbal short-term memory and vocabulary acquisition would therefore represent a significant theoretical advance. It is the primary aim of this paper to outline such a model.

TARGET PHENOMENA

The foregoing discussion raises the question of what would constitute an adequate computational account. A model that aims to account for the relationship between vocabulary acquisition and verbal short-term memory must at a minimum specify processing mechanisms that account for: (I) the ability to learn novel word forms after few exposures; (II) the ability to repeat novel word forms immediately; (III) production and repetition of known words; (IV) the key empirical results from the ISR paradigm; and (V) differences in immediate serial recall of lists of words vs nonwords. No model currently available can account for the full set of empirical results for each of these

basic human capacities. Let us examine each of these phenomena in greater detail.

Processing and Learning Words

Learning Novel Word Forms

The first phenomenon that must be accounted for by an integrated model is word learning. A number of studies have examined word learning in children with both normal and impaired abilities by presenting the children with multiple exposures to novel words and their referents and tracking the emergence of comprehension or production over time (Au & Glusman, 1990; Baldwin & Markman, 1989; Callanan & Markman, 1982; Carey, 1978; Carey & Bartlett, 1978; Dickinson, 1984; Dollaghan, 1985, 1987; Holdgrafer & Sorensen, 1984; Keefe et al., 1989; Leonard et al., 1981; Markman, 1984, 1989; Markman & Hutchinson, 1984; Merriman & Bowman, 1989; Merriman & Schuster, 1991; Mervis, 1984; Mervis & Pani, 1980; Mervis et al., 1994; Nelson & Bonvillian, 1973; Oviatt, 1980, 1982; Schwartz & Leonard, 1984). For example, using this approach, some studies have investigated the process whereby children extend the novel words to untrained referents (Nelson & Bonvillian, 1973; Oviatt, 1982), while others have examined the rate of acquisition as a function of the child's referential or nonreferential orientation or the phonological composition of the novel word (e.g., Leonard et al., 1981). All these studies have provided important documentation of various aspects of children's ability to acquire new words, but they have not been designed in ways that can elucidate the cognitive processes involved. That is, their findings advance knowledge about what kinds of word learning can and do occur, but do not directly further our understanding of the cognitive processing that underlies word learning.

In the studies most directly relevant to present purposes, it has been shown that normal preschoolers (ages 2 through 6) appear able to create "fast mappings" between a novel word form and its referent, with as little as one or two exposures to the new word (Carey, 1978; Carey & Bartlett, 1978; Dollaghan, 1985). Fast mapping occurs when the learner first notices a new lexical item and rapidly encodes it (i.e., creates a representation of the word form) after even a single experience with it. In natural settings, the learner will also store some semantic/conceptual information about it, creating a mapping between the form and the semantics.

Dollaghan (1985) found that 81% of her subjects were capable of lexical comprehension after a single exposure to the novel word and that 45% of her subjects were capable of making a recognizable attempt to say the word after only two exposures. The first exposure, it appears, moved children to what Keenan and MacWhinney (1987) have called a level of lexical comprehension, and the second exposure moved at least some of them to a level of comprehending imitation. These data are consistent with Keenan and Mac-

Whinney's (1987) observation that in the normal course of language acquisition children move quickly to at least a level of lexical comprehension of new words. They are also consistent with Carey's (1978) suggestion that, following their acquisition by fast mapping, phonological and semantic representations may be refined over a longer period of time.

These data on fast mapping therefore constitute the acquisitional phenomena that must be accounted for by a model: the ability to create representations of novel word forms within a very few presentations and to consolidate these representations over a period of time.

Immediate Repetition of Novel Word Forms

The second set of phenomena that must be accounted for by an integrated model relates to the repetition of novel forms. As previously discussed, nonword repetition is a measure of verbal short-term memory that has been used as an alternative to ISR. The fact that nonword repetition ability is highly correlated with vocabulary acquisition is thus an important part of the phenomena that have to be accounted for. The relationship seems intuitively plausible, because every new word a language learner acquires necessarily starts out, in effect, as a possible nonword (for the learner). Repetition of a nonword requires the formation of a temporary representation of it; learning a nonword requires development of a long-term representation of it. It therefore seems reasonable that superior nonword repetition abilities would be associated with superior vocabulary acquisition abilities.

As a starting point, therefore, the model must account for nonword repetition abilities. Normal human adults perform at close to ceiling in immediate repetition of nonwords. For example, on the Sound Mimicry subtest of the Goldman-Fristoe-Woodcock Sound-Symbol tests, which is a test of nonword repetition that provides norms for ages 3 through 80 (Goldman, Fristoe, & Woodcock, 1974), over 80% of subjects ages 19 through 27 made fewer than 5 nonword repetition errors in a test battery of 55 items. In testing repetition of a set of 40 nonwords, Gathercole, Willis, Emslie, and Baddeley (1991) have shown a clear developmental progression for the ages 4 through 6, with mean repetition scores increasing from 21.3/40 at 4 years to 30.7/40 at age 6. Nonword repetition becomes error-prone, however, with increasing nonword length: Gathercole et al. (1991) found consistent influences of nonword length on repetition accuracy at each of the ages 4 through 6, with repetition performance decreasing systematically as nonwords increased in length from two through four syllables.

These phenomena are therefore part of what must be captured by a model. As a first approximation, it seems reasonable to ignore the developmental progression of increase in repetition ability. What must still be accounted for is the excellent nonword repetition ability of both children and adults and the effects of nonword length.

Production and Repetition of Single Known Words

The third set of phenomena that must be accounted for by an integrated model relates to the processing of known words. Given that the model must account for the virtually error-free ability to repeat novel word forms, it would be strange if it did not offer an account of human speakers' ability to repeat known words. In doing this, the model must necessarily also provide some account of the spontaneous production of single words and of the recognition of known single words.

Verbal Short-Term Memory

Immediate Serial Recall

The fourth set of phenomena that must be accounted for by an integrated model relates to findings from the immediate serial recall paradigm.

Limits on span, the serial position curve, and chunking. Two basic features of immediate serial recall are that span size is limited, and that, even within this span, performance declines with increasing list length. Average span is around 7 items, as is well known (e.g., Miller, 1956). Recall is virtually perfect up to about 5 items and then declines dramatically to reach virtually zero performance as list length increases to 10 items (Guildford & Dallenbach, 1925). Any model of ISR must account for these data.

Short-term recall performance varies with serial position in the list of recall items. This so-called serial position curve, with its "primacy" and "recency" components, is one of the most characteristic features of short-term recall and also of ISR. The primacy effect refers to the advantage in recall of items in the first few positions in a list. The recency effect refers to the advantage in recall for the last few items on a list and is markedly greater with auditory than with visual presentation of the list. Items occupying the middle positions in a list are the least reliably recalled. These also constitute basic phenomena that a model of ISR must capture.

The essence of the articulatory loop hypothesis is that memory span will depend on the rate of rehearsal, with span being approximately the number of items that can be rehearsed in 2 sec. In this view, memory span is rate-limited. There is, however, the well-known finding that human memory span tends to reflect a constant number of chunks (7 ± 2 ; Miller, 1956), regardless of the characteristics of those chunks; this could be termed a more capacity-limited view of memory span. These two conceptions have been reconciled in the work of Zhang and Simon (1985), who showed that span is determined by rehearsal rate, in accordance with the articulatory loop view. However, rehearsal rate itself depends not only on articulation rate, but on the number of chunks (items) in the list to be recalled. This dependence of span on the number of items as well as articulation rate is an important part of the ISR phenomena that must be addressed. That is, any model should specify in

what way chunking and rehearsal processes underlie performance of the immediate serial recall task.

Phenomena suggesting an "articulatory loop." A number of phenomena have motivated and influenced thinking about the articulatory loop. It is the ability to explain these phenomena that is regarded as evidence of the model's success (e.g., Baddeley, 1990a).

The phonological similarity effect refers to the fact that, in immediate serial recall of lists of words, sequences of similar sounding words are recalled in correct order much less frequently than sequences of dissimilar words of comparable frequency and length. This result holds irrespective of whether the recall stimuli are presented in the auditory or the visual modality (Baddeley, 1986).

The irrelevant speech effect refers to the fact that immediate serial recall of lists of items is disrupted by the presentation of irrelevant spoken material not produced by the subject, despite the fact that the subject is free to ignore this material. The disruptive characteristics of the unattended material appear to be primarily phonological, with nonsense syllables being just as disruptive as meaningful words. Again, this effect obtains regardless of modality of presentation (Baddeley, 1986).

The word length effect refers to the finding that immediate serial recall performance of word sequences deteriorates as the constituent words in the sequence become longer, whether they are presented auditorily or visually (Baddeley et al., 1975).

The concurrent articulation effect refers to the finding that when the subject is required to engage in concurrent articulation, i.e., to articulate an irrelevant sound during list presentation, immediate serial recall is markedly impaired, under both visual and auditory presentation of stimuli (Baddeley, 1986).

Each of the first three effects has certain further interactions with the fourth. With auditory presentation, the phonological similarity effect is still observed under concurrent articulation; that is, when subjects perform ISR under concurrent articulation, they recall sequences of similar sounding words in correct order much less frequently than sequences of dissimilar words, just as in the phonological similarity effect without concurrent articulation. Similarly, under concurrent articulation with auditory presentation of recall stimuli, the irrelevant speech effect also seems to be somewhat preserved, at least under certain conditions (Hanley & Broadbent, 1987); that is, subjects' recall performance under concurrent articulation is further impaired by the presence of irrelevant speech. However, under concurrent articulation with auditory presentation, the word length effect is abolished if concurrent articulation is required during both list input and recall (Baddeley et al., 1984). That is, when subjects perform ISR under concurrent articulation, it no longer matters whether the words in the list are long or short. With visual presentation of recall stimuli, concurrent articulation abolishes the ef-

fects of phonological similarity, irrelevant speech, and word length (Baddeley, 1986).

Serial Recall for Words vs Nonwords

The fifth set of phenomena that must be accounted for by an integrated model relates to differences in the immediate serial recall of lists of words and list of nonwords. ISR performance is worse for lists of nonwords than for lists of known words (Hulme, Maughan, & Brown, 1991). Also, the types of errors made by subjects are different in serial recall of lists of familiar vs unfamiliar word forms (i.e., lists of words vs nonwords). In ISR of lists of known words, item order errors predominate, i.e., errors in which items in the list appear in the wrong position, often with interchanging of position (Aaronson, 1968; Bjork & Healy, 1974). In ISR of lists of nonwords, the predominant error type involves not the misordering of items in the list, but transposition of parts of words from one item in the list to another; these errors preserve the syllable structure of the target list and obey the phonotactic constraints of the language (Treiman & Danis, 1988).

As noted in discussing nonword repetition, every word known to a language learner was necessarily at one point an unknown word, i.e., a nonword. In fact, word acquisition can be viewed as the process by which a "nonword" becomes a "word" (for the learner); the relationship between words and nonwords is therefore important in understanding vocabulary acquisition. The present phenomena are important because accounting for them requires specification of how words and nonwords are represented, as well as specification of how these representations are accessed in ISR. These data are therefore of obvious relevance to any account of the relation between verbal short-term memory and vocabulary acquisition and so must be addressed by the model.

Summary

We now know a fair amount about various aspects of verbal short-term memory and about some aspects of word learning. How can the set of phenomena we have just outlined be captured in a computationally explicit model? There have been several attempts to model subsets of the phenomena. In the next section, we examine existing computational models of word learning, and in the following section we examine existing models of verbal short-term memory. Our discussion of each of these formulations is with an eye not only toward describing the model, but also toward specifying the important insights they express, as well as any major limitations they reflect.

COMPUTATIONAL MODELS OF WORD LEARNING

Learning Serial Order (Grossberg, 1978)

Grossberg (1969, 1978, 1986) has presented numerous ideas pertaining to serial order that are relevant to word learning as well as to aspects of short-

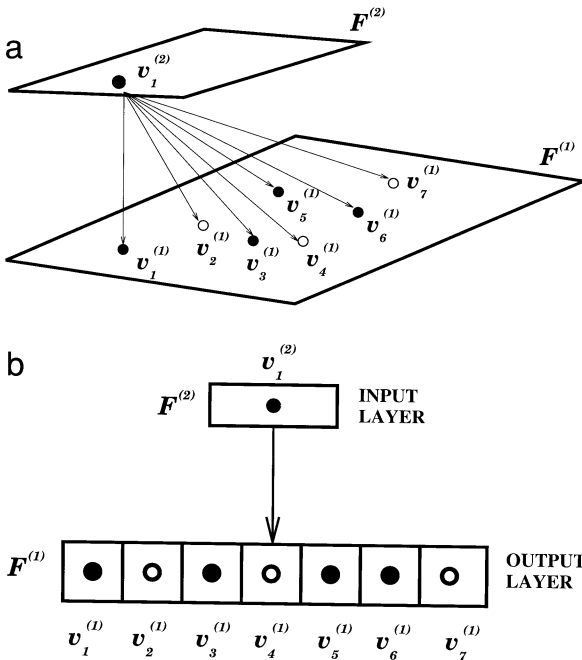


FIG. 1. Outstar learning. (a) Typical depiction of an outstar. (b) Equivalent two-layer connectionist network.

term memory. However, the ideas have not been developed explicitly enough to constitute a model of word learning.¹ They have also not been developed as an explicit account of immediate serial recall phenomena. The following discussion summarizes some of these ideas very briefly.

Outstar Learning of Spatial Patterns

A field is a population of nodes (cells) that constitutes a level of representation and corresponds to a “layer” of connectionist units. In Grossberg’s notation, two fields might be denoted as $\mathfrak{S}^{(1)}$ and $\mathfrak{S}^{(2)}$, and cells within these fields would be denoted as $v_1^{(1)}, v_2^{(1)}, v_3^{(1)}, \dots$ and $v_1^{(2)}, v_2^{(2)}, v_3^{(2)}, \dots$, respectively.

An outstar is a cell in one field that “samples” signals that are active over another field. For example, cell $v_1^{(2)}$ in field $\mathfrak{S}^{(2)}$ might sample the activity of cells $v_1^{(1)}, v_2^{(1)}, v_3^{(1)}, \dots, v_i^{(1)}$ in field $\mathfrak{S}^{(1)}$. This is depicted in Fig. 1a. Cell $v_1^{(2)}$, together with its projections that gather input from $\mathfrak{S}^{(1)}$, is an “outstar.” An outstar learns a spatial pattern over its sampling domain, by Hebbian adjustment of its projection weights. In more familiar terms, an outstar is

¹ Adaptive Resonance Theory (ART; Grossberg, 1987) is also relevant to issues in word learning, especially to the question of how a system can remain plastic without catastrophic interference to existing representations. However, ART does not deal with serial order.

equivalent to a two-layer connectionist network composed of a single unit ($v_1^{(2)}$) at the input layer and a number of units at the output layer ($v_1^{(1)}, v_2^{(1)}, v_3^{(1)}, \dots, v_i^{(1)}$). Figure 1b shows this connectionist formulation. Comparison of the two figures should make it clear how the two formulations are equivalent. In the two-layer network, Hebbian adjustment of weights from the single-unit input layer to the output layer can enable subsequent activation of the single input-layer cell to reproduce the previously "sampled" pattern of activation over the output layer. Thus, outstar cells in $\mathfrak{S}^{(2)}$ can learn to recreate patterns of activation over $\mathfrak{S}^{(1)}$. This is outstar learning.

Learning Sequences with an Avalanche

An avalanche is an assembly consisting of a cell (which we will call the "avalanche node") that has uniform connections to a set of outstars $v_i^{(2)}$ in field $\mathfrak{S}^{(2)}$. When the avalanche node is activated, the outstars become activated in sequence. This is depicted in Fig. 2a, which shows the avalanche at several time steps. For each time step, the top-most, active node is the avalanche node. Below it are arrayed the outstars in field $\mathfrak{S}^{(2)}$. The figure shows that outstars $v_1^{(2)}, v_2^{(2)}, v_3^{(2)}, \dots$ (in field $\mathfrak{S}^{(2)}$) are activated successively at time steps 1, 2, 3, . . . At any one time step, only one outstar is active.

Figure 2b shows how the avalanche can encode serial order. The figure depicts several time steps of processing. The top-most node shown in the illustration for each time step is the avalanche node. Below it are the outstars in field $\mathfrak{S}^{(2)}$. The outstars each sample activity over field $\mathfrak{S}^{(1)}$. Exactly one outstar $v_i^{(2)}$ samples activity over $\mathfrak{S}^{(1)}$ at any particular time step t .

Consider the situation when cells $v_1^{(1)}, v_2^{(1)}, v_3^{(1)}$, and $v_4^{(1)}$ in $\mathfrak{S}^{(1)}$ are sequentially activated by a series of inputs, at time steps 1, 2, 3, 4. Assume that the avalanche node is activated at time step 1. Then, outstars $v_1^{(2)}, v_2^{(2)}, v_3^{(2)}, v_4^{(2)}$ will be activated in succession at time steps 1, 2, 3, 4, and each one will sample the pattern of activity over $\mathfrak{S}^{(1)}$ at exactly one time step.

This is depicted graphically in Fig. 2b, which shows several time steps of processing. At each time step, exactly one outstar is active in $\mathfrak{S}^{(2)}$, and exactly one element is active at $\mathfrak{S}^{(1)}$. As before, the sequence of activations at $\mathfrak{S}^{(1)}$ represents the serial ordering that is to be learned. Figure 2b shows that each $\mathfrak{S}^{(2)}$ outstar will sample exactly only one spatial pattern over $\mathfrak{S}^{(1)}$. For example, at time step $t = 1$, outstar $v_1^{(2)}$ is active and therefore samples the activation pattern at $\mathfrak{S}^{(1)}$, which in this case consists of the activation of cell $v_1^{(1)}$. Hebbian adjustment of weights from $v_1^{(2)}$ to $v_1^{(1)}$ will therefore take place. A similar process occurs at each successive time step, each time involving a different outstar.

If the avalanche node is activated subsequently (in the absence of input at $\mathfrak{S}^{(1)}$), the outstars $v_1^{(2)}, v_2^{(2)}, v_3^{(2)}, v_4^{(2)}$ will once again be activated in succession. Because of the Hebbian weights on the connections from each of these outstars to $\mathfrak{S}^{(1)}$, activation of each outstar will result in reinstatement over

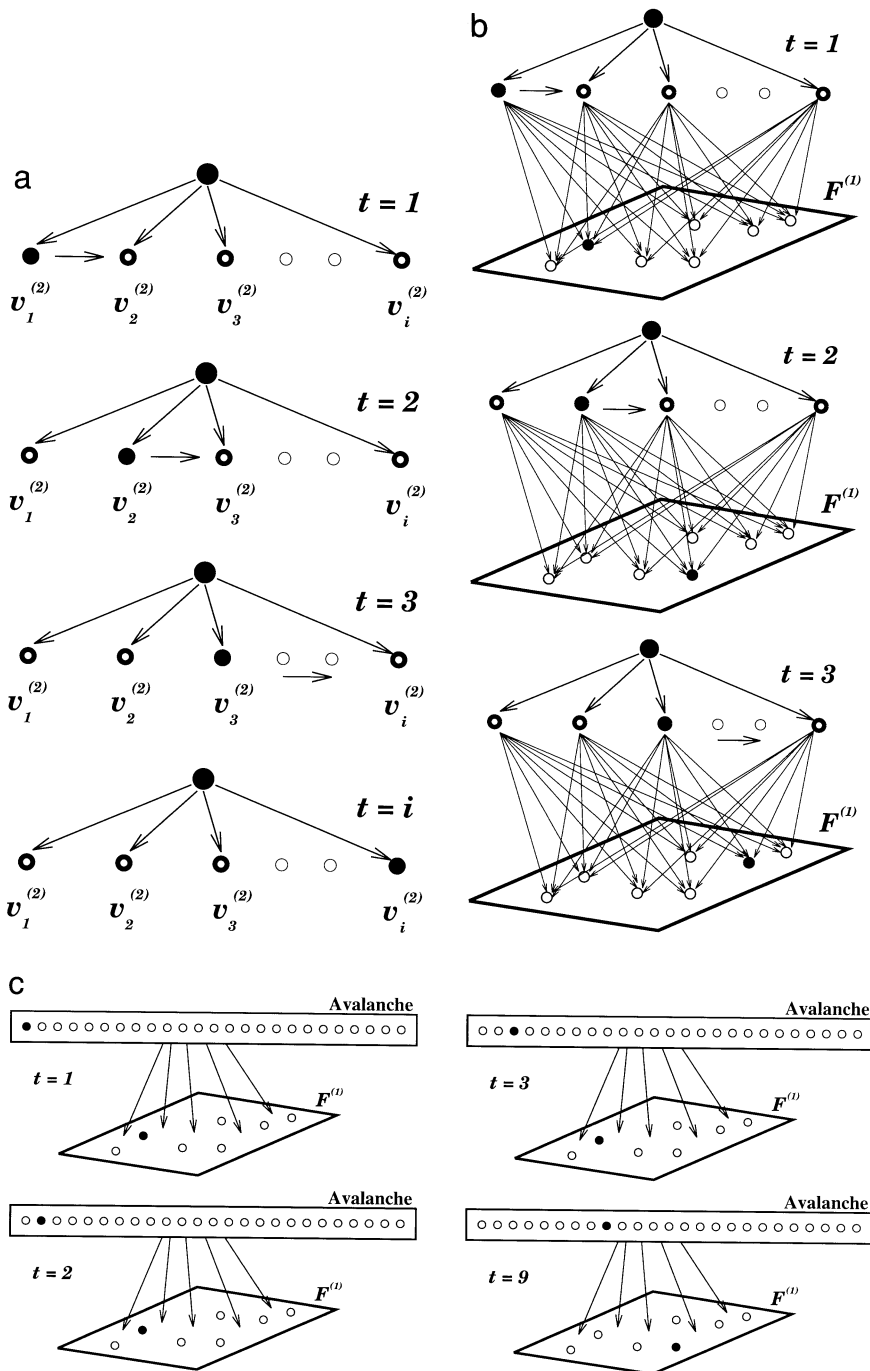


FIG. 2. Encoding serial order with an avalanche. (a) Successive activation of outstars in an avalanche. (b) Encoding of a serial pattern by avalanche outstars. (c) Abstract depiction of how an avalanche encodes serial order.

$\mathfrak{S}^{(1)}$ of the pattern of activation that was present during learning. Thus, the sequence of activations $v_1^{(1)}, v_2^{(1)}, v_3^{(1)}, v_4^{(1)}$ will be recreated over $\mathfrak{S}^{(1)}$.

If we distill the key ideas underlying the avalanche, we arrive at the depiction in Fig. 2c. This shows that the avalanche of outstars in field $\mathfrak{S}^{(2)}$ can be treated abstractly as elements in an array. The elements in this array sample activations over field $\mathfrak{S}^{(1)}$. Activation of the avalanche amounts to a wave of activation traveling in a fixed sequence along the $\mathfrak{S}^{(2)}$ array, activating exactly one array element at a time. At each time point, the connections from the currently active array element to field $\mathfrak{S}^{(1)}$ are adjusted by a Hebbian process. This means that each $\mathfrak{S}^{(2)}$ array element will encode whatever pattern of activation was present over $\mathfrak{S}^{(1)}$ at that time step. The process of recall requires that the wave of activation must travel along the array once again. When it does, each array element will recreate its encoded pattern of activations over $\mathfrak{S}^{(1)}$.

Recall that an outstar is a two-layer network with a single input unit. When learning a sequence, this network has to associate the sequence of decaying activation values of the single input unit with a sequence of output patterns. An avalanche is a two-layer network with n inputs, each representing an outstar. It amounts to a way of increasing the orthogonality of input vectors, thus facilitating the Hebbian learning of output sequences.

Competitive Queueing (Houghton, 1990)

Houghton (1990, 1993, 1994) has developed a computational architecture called competitive queueing to address issues involved in memory for sequences of items. The competitive queueing (CQ) model is very similar to Grossberg's ideas described above and aims to provide mechanisms by which:

1. At any time, several computational elements (nodes, or units, each representing some conceptual entity, such as a word, a syllable, or a proposition) intended for sequential performance can be simultaneously active.

The purpose of this is to provide a basis for serial order errors, which are ubiquitous in sequential domains and involve not only anticipations ("A, B, C" \rightarrow "A, C, C"), but transpositions ("A, B, C" \rightarrow "A, C, B").

2. Serial order information is available, whereby an unambiguous sequence can be constructed from these multiple active elements.

3. Sequences presented to the system, and represented in the above manner, can be learned from very few presentations and subsequently reproduced from memory.

The mechanisms in the CQ model that actually provide this functionality are as follows (e.g., Houghton, 1993):

- Nodes representing the elements composing the sequence are activated one at a time in a network of nodes ("Layer 1"). Their activation represents the sequential presentation of items in the stimulus sequence. The activation

of each node decays following its point of initial activation. Thus multiple Layer 1 nodes will be active, at all time points except the first.

- The activations of multiple nodes over Layer 1 constitute an activation gradient, which represents the serial order information. Note that this gradient is established by input of the stimulus items themselves, together with decay.

When a sequence is presented to the system, the most recently activated element in the sequence will have the highest activation of all elements (because previously activated nodes in the sequence will have decayed). Consequently, the activation gradient is a recency gradient.

- To provide for learning of sequences, the CQ model posits a bank of higher-level nodes (Layer 2). Each of these Layer 2 nodes emits a context signal that changes at each time step. This context signal consists of a pattern of activation over two outstar nodes, which are designated as a start node and an end node. Each of these outstars has weighted connections to nodes in the network over which the sequence is represented (Layer 1).

At the beginning of presentation of a sequence of elements over Layer 1, a Layer 2 node is recruited. Its start node is activated, and the start node's activation decays throughout presentation of the sequence of stimuli. At each time step during presentation, weights from this start node to Layer 1 are adjusted via Hebbian learning. This outstar's weights encode a primacy gradient. At the end of sequence presentation, the Layer 2 node's end node is activated. Its weights are adjusted and encode a recency gradient.

During recall, the same Layer 2 node is activated. Its start node is also activated. Activation of the start node decays, while activation of the end node increases. This is arranged in such a way that the sum of activations of the start and end nodes is a constant. Thus at the beginning of recall, the start node has much higher activation than the end node and therefore controls activation delivered to Layer 1 nodes. As noted above, this imposes a primacy gradient on Layer 1. Toward the end of recall, the end node will be more active than the start node and will therefore tend to impose a primacy gradient on Layer 1. Thus the start and end nodes complement each other in delivering input to the Layer 1 nodes.

- The CQ model incorporates a "competitive filter" associated with Layer 1. This is a winner-take-all network that allows exactly one Layer 1 node to fire at each time step, by choosing the node with the highest activation. The competitive filter then inhibits the just-fired node, so that the node with the next-highest activation will be chosen next. During recall, the competitive filter provides for response selection at Layer 1, at each time step.

The CQ model is in many ways equivalent to the computational formalisms developed by Grossberg (1969, 1978, 1986). The representation of temporal information by an activation gradient is a key idea in Grossberg's work and is also adopted in the CQ model. In the CQ model, each Layer 2 node

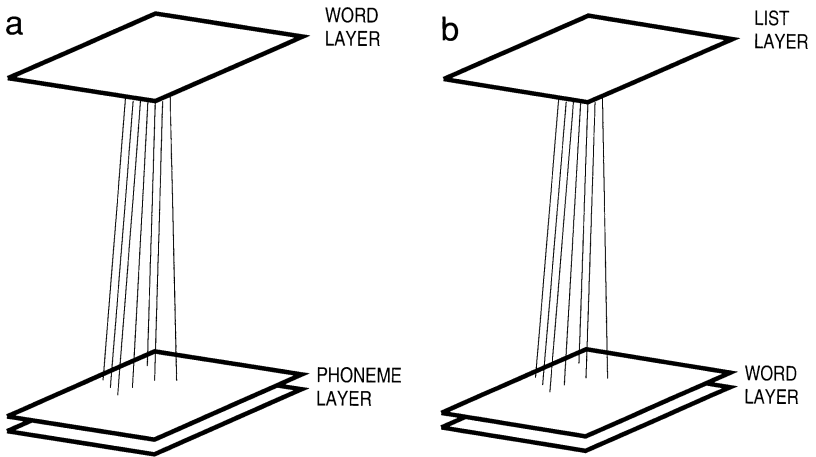


FIG. 3. (a) Competitive Queueing model of word learning (Houghton, 1990). (b) Model of immediate serial recall of lists of words (Burgess & Hitch, 1992).

encodes serial order through a combination of outstar sequence learning and the avalanche sequence learning discussed in the review of Grossberg's ideas in the previous section. The notion of a competitive filter in the CQ model is equivalent to the self-inhibition of ordered STM traces in Grossberg's (1978, 1986) framework.

Even though Grossberg's work represents the earlier formulation of these ideas, it is the CQ model that will be used as the reference point in the following review of computational models, for two reasons. First, the CQ framework states the key ideas in a way that is somewhat more accessible for present purposes and is therefore easier to work with. Second, a number of the existing, implemented computer models have taken Competitive Queueing as their base and so are more easily discussed using CQ terminology.

If a word form is viewed as a sequence of elements (such as phonemes), then the CQ model, like Grossberg's work, provides many of the elements of an account of fast mapping. This is exactly the approach taken in Houghton (1990), which applies the CQ framework to the learning of new words. The model is depicted in its essentials in Fig. 3a. One layer of nodes represent phonemes (the "phoneme layer"). A second layer of the system (the "word layer") contains nodes representing words. When a new word is presented to the system (phoneme by phoneme), the appropriate phoneme nodes are activated in sequence in the phoneme layer. Activations of these nodes decay, so that an activation gradient is established, in which the most highly activated node represents the most recently presented phoneme. However, multiple phonemes will be simultaneously activated, at various levels

of activation. Associated with the phoneme layer is a competitive filter ensuring that only one phoneme at a time controls output. Nodes in the competitive filter are in one-to-one correspondence with nodes in the phoneme layer; there is a hard-wired excitatory connection from each phoneme layer node to one particular competitive filter node and a reciprocal hard-wired inhibitory connection from the competitive filter node to its phoneme node. At any time step, the most active phoneme node at the phoneme layer will cause its corresponding competitive filter node to win the competition, and thus this will be the only phoneme allowed to produce an output at that time step. The competitive filter node then strongly inhibits its corresponding phoneme layer node, so that at the next time step, the phoneme layer node with the next highest activation will control output, and so on. In this way, the sequence of input phoneme activations can be immediately reproduced as a sequence of output activations.

At the beginning of presentation of a new word, a word node is recruited in the word layer; the temporal signal it emits (see discussion of CQ models above) is a two-element vector. The activation pattern over this two-unit mini-layer changes from [1 0] to [0 1] over the time course of presentation of the word. At each time step, i.e., when each phoneme is presented, weights from this word layer vector to the phoneme layer are adjusted by Hebbian learning. In this way, one-shot learning of the new word takes place. Subsequent activation of that particular word node will result in the temporal signal being emitted, which in turn will result in output of the phoneme sequence constituting that word.

Clearly, the Houghton (1990) model directly addresses issues involved in the learning of new words, in particular, as regards fast mapping. It also exhibits word length effects, in that recall of words declines with increasing word length, i.e., number of phonemes. However, it is only a partial account, formulated to deal only with monosyllabic words: words longer than seven phonemes in length are very poorly learned. Seven phonemes is a reasonable length limit for monosyllabic words, but it is insufficient if polysyllabic words are to be taken into consideration.

The model clearly also accounts for the immediate repetition of monosyllabic words and nonwords. Again, however, polysyllabic words/nonwords are beyond the scope of the model.

A further limitation of the model is that there is no means of differentiating between known and novel phonological sequences. As formulated in Houghton (1990), the model would allocate a new node to such a word and learn it all over again as a new word. The model therefore does not address issues regarding the distinction between words and nonwords.

Finally, the model does not address issues related to verbal short-term memory for lists of items. As a result, the model does not offer an account of the relationship between vocabulary acquisition and verbal short-term memory.

Acquiring a Distributed Lexicon (Miikkulainen, 1990)

Miikkulainen (1990) has developed a distributed feature map model of the lexicon (DISLEX) which includes a model of word learning. DISLEX is built around self-organizing topological maps (Kohonen, 1984). A self-organizing network or map consists of a pool of units that have two-dimensional spatial structure. The spatial structure is interpreted to mean that two neighboring units in the map represent more similar information than do two units that distant from each other. For example, DISLEX incorporates one such map (the physical map) that represents visual word forms. The representation of any particular written word, e.g., *dog* is a particular distributed pattern of activation over the physical map. These representations are formed by a topology-preserving unsupervised learning process in which n -dimensional representations of the visual features of word forms are projected onto the two-dimensional physical map (Kohonen, 1984). Thus the topology of the resultant feature map reflects visual similarities among words. Stated another way, if the physical map representations of two words are similar, this will be because the two words are visually similar.

A second map in Miikkulainen's (1990) model is a semantic map, also formed by a topology-preserving unsupervised learning process in which n -dimensional representations of the semantic features of words are projected onto the two-dimensional semantic map. Thus the semantic map's activity pattern will be similar for two words that have similar meanings.

Bidirectional connections are also formed between these two maps, by Hebbian weight adjustment. As a result, activation patterns over the physical map can evoke the corresponding pattern over the semantic map and vice versa. That is, presentation of a visual word form can evoke the word's semantics, and presentation of a word's semantics can evoke the visual word form (Miikkulainen, 1990).

This constitutes DISLEX's account of word-learning. The topological feature map could easily be a model of phonological feature space instead of visual word form feature space. It would then offer an account of the learning of phonological forms and their mapping onto a semantics.

This account of word learning has the advantage of providing distributed representations for words and of linking word form with semantics. However, it fails to capture the fact that words have temporal extent, instead treating words as occurring instantaneously. Furthermore, it is not formulated so as to address verbal short-term memory phenomena. It therefore does not offer the required integrated account.

COMPUTATIONAL MODELS OF VERBAL SHORT-TERM MEMORY

ISR of Lists of Words (Burgess & Hitch, 1992)

Burgess and Hitch (1992) use the CQ framework to develop a model of Immediate Serial Recall, depicted in Fig. 3b. Their model has a layer of

nodes representing words [corresponding to the layer representing phonemes in the Houghton (1990) model]. A particular word node can be activated by an input vector representing the phonemes in the word. For example, an input vector encoding the phoneme string /s æ t/ will activate the word node for *sat*; note that a word is treated as having no temporal dimension. Weights enabling a particular input vector to activate a particular word node in the word layer are assumed to have been set by external processes and are fixed and permanent for the purposes of the model.

Another layer of nodes (the "list layer") represents elements to which a sequence of words can be temporally bound. These list layer nodes emit a temporally varying signal. When a list of words is presented to the system (at the word layer), the various word nodes are activated, establishing an activation gradient, in standard CQ fashion. A list layer node is recruited at the beginning of presentation; at each time step, weights from the current state of its temporal signal to the word layer are adjusted via Hebbian learning. A competitive filter associated with the word layer provides for selection and output of single words from the pool of multiply active word nodes.

The close structural correspondence with the Houghton (1990) model is worth noting explicitly. The Burgess and Hitch (1992) list layer and word layer correspond respectively to the word layer and phoneme layer in the Houghton (1990) model (compare Figs. 3a and 3b). Just as a sequence of phonemes can be bound to a word layer node in the Houghton (1990) model, a sequence of words can be bound to a list layer node in the Burgess and Hitch (1992) model. The mechanisms are essentially identical: Hebbian learning binds a sequence of nodes at one layer to a single node at another layer, using a temporal signal emitted by the higher-layer node. However, because recall of a sequence is meant to be temporary, the Burgess and Hitch (1992) model adds decay to the weights established by Hebbian learning between the list layer node and word layer nodes; the model also adds noise to the activations of nodes. This basic model, together with a few additional features discussed below, enables simulation of a wide range of verbal short-term memory phenomena.

The model exhibits basic limitations on its memory span; this is because temporary weights undergo more decay by the end of list presentation if there are more items in the list. Thus there is poorer recall for longer lists of words. Two aspects of the model's functioning tend to provide a basis for serial position effects. First, the overlap between the context vectors at various time steps is least at the beginning and end of lists. That is, the context vectors at the initial and final time steps are most discriminable from all the other context vectors; the context vectors at middle time steps have more overlap with each other. For this reason, the initial and final positions of the list will tend to be better recalled. Second, if an error has been made in recall, i.e., the wrong word node was activated at a particular serial posi-

tion, the chance of further errors keeps increasing down the list; the first serial position is immune to this, while the last serial position is most prone to this. These two factors together lead to the model exhibiting a strong primacy effect. The two factors cancel each other out, however, with respect to the final serial position, and so the model exhibits no recency effect. However, this is in part due to an extraneous "chaining" mechanism used in the model, which the authors themselves advocate abandoning at the end of their paper. This chaining mechanism operates so as to establish links (connection weights) between one list item and the next, during presentation. During recall, output of a particular list item therefore cues the next item via these chaining weights.

The model captures word length effects by assuming that the decay of weights occurs once per phoneme rather than once per word; the greater the number of phonemes in a list, the greater the decay. This allows for word length effects, as a list of n words of average length k phonemes will be subject to less weight decay than a list of n words of average length greater than k phonemes. This also allows the model to address the relationship between the capacity-limited and rate-limited views of span. Decay (and hence span) is related to the number of phonemes in the list; however, the number of words (i.e., chunks) will also be relevant to determining span, because the greater the number of excited word nodes, the greater the possibility of noise-induced errors.

The model captures the phonological similarity effect because input vectors produce activation of not only the target word, but also partial activation of other similar words. When the input vectors for a sequence of words are similar to each other, each will contribute to the activation of the other similar words in the list, and so the activation gradient will tend to be reduced, and the possibility of noise-induced errors will increase. The model is also successful in simulating errors in serial recall, in that it produces more order errors than item errors and in that phonemic similarity has its effect chiefly on order errors.

However, the account of "rehearsal" in the model is not very clear. The authors appear to interpret rehearsal as consisting of the use of the chaining mechanism. This is a questionable assumption to begin with. Later, the authors recommend dispensing with the chaining mechanism altogether (for other reasons), but it is unclear what rehearsal would then consist of. The rehearsal mechanism is therefore either problematic or does not exist, depending on what version of the model is considered. This criticism also applies to the model's account of the effects of concurrent articulation, which is assumed to affect use of the chaining mechanism.

A further limitation is that word length effects and chunking effects are both achieved by having weights decay once per phoneme rather than per word, which is intended to simulate the time it takes to articulate the word. However, the assumption that it takes time to articulate a word seems incon-

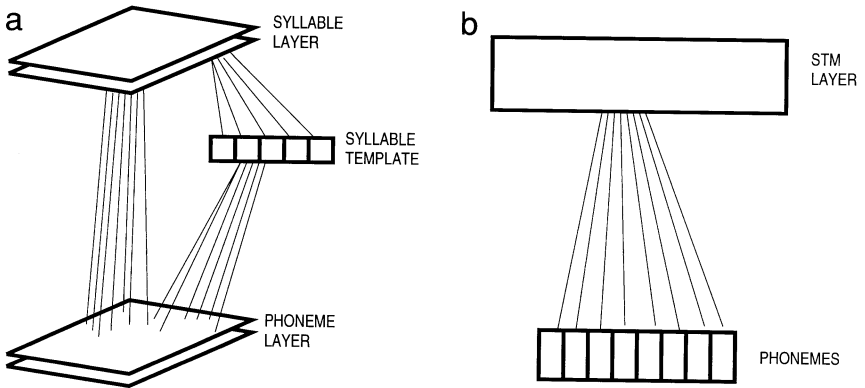


FIG. 4. (a) Model of immediate serial recall of lists of nonwords (Hartley & Houghton, 1996). (b) Model of immediate serial recall of lists of words and nonwords (Bairaktaris & Stenning, 1992).

sistent with the model's treatment of a word as having no temporal duration for the purposes of recognition.

Most of the word-level phenomena we are currently interested in simulating fall outside the scope of the model. The model does not address the processing of nonwords, either for immediate repetition or for serial recall of lists of nonwords. This of course means that the model does not address the learning of novel words or processing differences between words and nonwords. This in turn means the model does not offer an account of the relationship between vocabulary acquisition and verbal short-term memory.

In summary, the Burgess and Hitch (1992) model succeeds in covering most of the verbal short-term memory phenomena it sets out to address, but it does not provide the necessary integrated account of vocabulary acquisition and verbal short-term memory.

Recently, Burgess (1995) has extended the Burgess and Hitch (1992) model to provide an account of how word forms can be learned as spatial patterns of activation. While potentially promising, this approach still has the limitation that it treats words as having no temporal duration. Moreover, the model does not account for the error types observed in human serial recall of lists of nonwords.

ISR of Lists of Nonwords (Hartley & Houghton, 1996)

A model proposed by Hartley and Houghton (1996) addresses repetition and immediate serial recall of nonwords, again using the CQ framework. The model is depicted in Fig. 4a. As in the Houghton (1990) model of word learning, there is a phoneme layer. Word forms (here, nonwords) are represented phoneme by phoneme, and, as in other CQ models, a gradient of phoneme node activations is established, representing the word form. The

higher-order layer to which this sequence is bound is a syllable layer. Each syllable node has a two-bit context signal, the two bits representing onset and rime, respectively.

To this standard CQ structure Hartley and Houghton add a further layer of structure. In addition to the layers just described, there is a "syllable template," which can be thought of as a specialized higher-order CQ layer consisting of just one node. This node has an associated five-bit context signal representing a simplified version of syllable structure: the elements of the vector represent two slots for prevocalic consonantal segments (syllable onset), one slot for a vowel (syllable nucleus), and two slots for post-vocalic consonantal segments. One cycle of activations of this context signal ($\{10000, 01000, 00100, 00010, 00001\}$) represents cycling through the structure of one syllable. This structure is assumed to be inbuilt hardware expressing constraints on articulation. Each element of the context signal has connections to those phoneme nodes in the phoneme layer representing phonemes that can legally occupy that position in a syllable, in a particular language. That is, weights from the syllable template to the phoneme layer encode the phonotactic constraints of the language. These are assumed to have been acquired by previous experience; from the point of view of the model, they are permanent weights.

When a sequence of phonemes representing a nonword is presented to the system, elements of the syllable template are activated in sequence, in accordance with which component of the syllable the currently most active phoneme can belong to; the syllable template thus acts as a parser.

Thus there are two sequences of input to be bound to a syllable layer node: the activation sequence at the phoneme layer and the activation sequence of the syllable template. Both of these are bound by Hebbian adjustment of weights from the syllable node to the syllable template and from the syllable node to the phoneme layer. When this occurs, the system has learned both the structure of a syllable and its constituent phonemes. Subsequent activation of the syllable node enables recall of the syllable. As the focus is on immediate serial recall of nonwords, the model assumes decay of the Hebbian weights from syllable nodes to the syllable template and to the phoneme layer; that is, the bindings of phonemes to syllable nodes are temporary.

The Hartley and Houghton model combines elements of the Houghton (1990) word learning model and the Burgess and Hitch (1992) model of immediate serial recall of words. As in the Houghton model, sequences of phonemes are bound to a node representing a monosyllable. Unlike the Houghton model, however, these bindings are meant only for temporary immediate serial recall and therefore do not result in long-term learning of these syllables. As in the Burgess and Hitch model, nodes at the syllable level themselves need to be ordered for serial recall. Unlike the Burgess and Hitch model, the Hartley and Houghton model does not implement this serial recall mechanism. That is, Hartley and Houghton assume the existence of a further

CQ mechanism that binds syllable nodes together into a list, as in the Burgess and Hitch (1992) model, but do not implement it. Hartley and Houghton simply assume that syllable node activations occur in the correct order. This is because the main focus of their model is on examining the error patterns of migration of phonemes from one syllable to another, under the assumption that the syllable nodes themselves are correctly ordered.

In serial recall of lists of words, the predominant errors are item errors, in which items in the list are transposed or otherwise wrongly ordered. In serial recall of lists of nonwords, by contrast, the human data exhibit errors that interchange parts of word forms, but that adhere to the phonotactic constraints of the speakers' language. The motivation for introduction of the syllable template in the Hartley and Houghton model is to be able to capture this pattern of nonword errors.

The model does capture nonword repetition abilities, exhibiting performance similar to that observed in nonword repetition by human adults, and also a similar pattern of errors. The model also exhibits the same types of interchange errors in serial recall of nonword lists as observed in the human data. The model's recall performance declines with list length, exhibiting limitations on memory span for nonwords as well as a small recency effect, although no primacy effect is shown.

A limitation of the model is that it deals with lists of syllables. A nonword presented for repetition is treated simply as a list of syllables; such a nonword is in no way different from a list of syllables presented for serial recall. That is, there is no notion of "word" in the model. Accordingly, the model does not address any of the differences in repetition of words vs nonwords, nor the difference in error patterns in serial recall of words vs nonwords. The model also does not address serial recall of lists of words or any of the associated phenomena. Also, there is no account of the learning of new words.

The authors do suggest adding a second syllable layer for words, exactly similar to the existing one for nonwords, but in which weights would not decay. This seems somewhat ad hoc, however, and still does not provide any notion of "wordness" for either real words or nonwords, i.e., everything is still a list of syllables. The authors also do suggest that initially temporary weights could become permanent through rehearsal (thereby providing a means of learning new words), but this idea is not developed any further.

In sum, the model does very well in its assigned task of simulating the error types observed in serial recall of lists of nonwords, but does not provide an account of the relationship between vocabulary acquisition and verbal short-term memory.

Other Models

The Bairaktaris and Stenning (1992) model is depicted in Fig. 4b, which shows that a word/nonword stimulus is presented to the system as a vector of constituent phonemes. Each time such a stimulus is presented to the system, a

node is allocated in the "STM layer," and weights from the STM layer node to the phoneme layer are adjusted by Hebbian learning. During list presentation, firing thresholds are set for these allocated STM nodes in such a way as to reflect the order in which list items are being encountered. During subsequent recall, these thresholds lead to the STM nodes firing in the original order, thereby simulating serial recall.

As an account of verbal short-term memory, this model does not appear to improve on the Burgess and Hitch (1992) model, nor does it address the issue of the learning of words. Like the other models considered above, therefore, it does not provide an account of the relationship between vocabulary acquisition and verbal short-term memory.

A model of immediate serial recall developed by D. W. Glasspool (manuscript in preparation) extends the Burgess and Hitch (1992) model to ISR of lists of nonwords. The model employs two parallel CQ mechanisms, one for word forms, treated as sequences of phonemes, and one for lists, treated as sequences of word forms, as in the Burgess and Hitch model. However, as noted by Hartley and Houghton (1996), this approach predicts that the same kinds of paired transposition errors will occur in ISR of list of nonwords as in ISR of lists of words; but in fact, as discussed in the previous section of this paper, error types differ in these two kinds of lists. For this and other reasons, the Glasspool model also does not offer a comprehensive account of the target phenomena.

A number of other computational approaches have examined aspects of short-term memory and/or issues relating to the maintenance of serial order (e.g., Amit, Sagi, & Usher, 1990; Elman, 1990; Jordan, 1986; McNaughton & Morris, 1987; Wang & Arbib, 1991). However, none of these models directly addresses the psychological data in immediate serial recall, and thus will not be reviewed here.

INTEGRATING VOCABULARY ACQUISITION AND VERBAL SHORT-TERM MEMORY: TOWARD A UNIFIED ACCOUNT

Clearly, none of the existing computational models offers a comprehensive account of the range of phenomena that need to be addressed in integrating vocabulary acquisition and verbal short-term memory. Each model covers some of the data but ignores other aspects. Nevertheless, each has important insights to offer.

The model proposed below represents the first comprehensive attempt at the important task of integrating the two fields of vocabulary acquisition and verbal short-term memory and addresses the range of data outlined in the second section of this paper. It attempts to incorporate the strengths of existing models wherever possible, while extending coverage to achieve the necessary unified account. It is worth emphasizing again that, even though the model is not currently implemented, it builds on existing work and there-

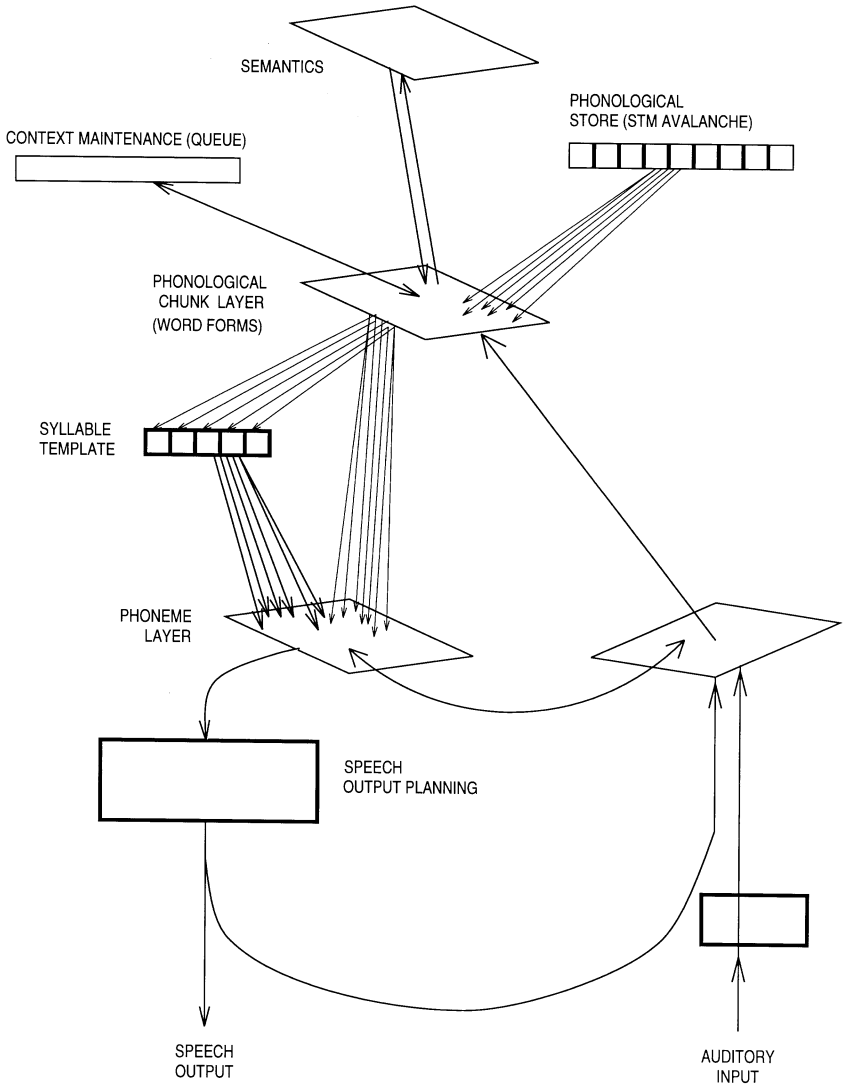


FIG. 5. Proposed model of vocabulary acquisition and verbal short-term memory.

fore provides a viable framework. In particular, the starting points for the present proposal are the models developed by Hartley and Houghton (1996) and by Burgess and Hitch (1992).

Outline of the Model

The model is shown in Fig. 5. The phoneme layer and phonological chunk layer are CQ structures of the kind already discussed above: each comprises

a set of nodes, together with an associated competitive filter. Nodes at the phoneme layer represent individual phonemes. Nodes at the phonological chunk layer represent "chunks," i.e., groupings of one or more syllables. For example, a monosyllabic form such as the noun *dog* and a polysyllabic form such as the noun *unwholesome* are both considered to be phonological chunks and would each be represented by a node at the chunk layer. This layer is assumed to be topologically organized; thus, phonologically similar forms such as *bat* and *cat* would have similar (neighboring) representations.

The phonological store is the same as what we termed the "list layer" in the Burgess and Hitch (1992) model. It is an avalanche of the kind described earlier, and its outstars sample activity over the phonological chunk layer. There is a syllable template between the phoneme layer and chunk layer. This functions exactly as in the Hartley and Houghton model and, as in that model, is interpreted as reflecting constraints on articulatory hardware.

Semantics is depicted as sending input to and receiving input from the phonological chunk layer. This adopts a view in which lexical organization is conceived of as a collection of topologically organized maps, as in our earlier discussion of the DISLEX model (Miikkulainen, 1990), which demonstrates how such a distributed lexicon could be formed by Hebbian learning. What is important for present purposes is that chunk layer nodes receive activation from their associated semantics at the semantics layer and, conversely, can activate their associated semantic representations at the semantics layer.

A context maintenance system (or queue) associated with the chunk layer stores sequences of activity patterns over the chunk layer, that is, it encodes a temporal sequence of vectors each of which represents the pattern of activation over the chunk layer at one point in time; at each time point, there will be one chunk substantially more active than the others. In this way, the queue encodes memory of a sequence of chunks. The queue operates along the lines of the Bairaktaris and Stenning (1992) model. Note that the phonological store and the queue both provide for encoding of sequences of spatial patterns over the chunk layer. The difference is that such encoding in the phonological store is an automatic process, whereas encoding in the queue is a controlled, strategic process.

The speech output planning system comprises mechanisms necessary for the motor programming of speech. In the present model, the existence of these is simply assumed. Input to speech planning is assumed to come from phoneme layer activations, possibly with intervening transductions of information. The outputs of speech planning lead to actual speech output. However, via an "efference copy," they are also assumed to activate the phoneme and chunk layers in the same way that auditory input does.

The topological organization of the chunk layer could come about through the formation of a self-organizing map (Kohonen, 1984) at the chunk

layer, in response to input of activation patterns from the phoneme layer. However, we do not describe such processes and simply assume their existence, and assume that they result in the topological organization of the chunk layer.

The model further assumes the existence of processes whereby word recognition occurs. In particular, we assume a model of speech perception something like TRACE (McClelland & Elman, 1986). The TRACE model has sets of detectors at three levels: the phonetic feature level, the phoneme level, and the word level. We assume that this TRACE-like speech perception front-end delivers to our present model an ordered set of phonemes and an ordered set of words. This input causes activation of one Phoneme Layer node at a time and one Chunk Layer node at a time, at each time point in processing, during presentation of word forms.

When a known word is presented to our system, four things are assumed to occur. First, the phonemes composing the word impinge on the phoneme layer one by one (following their identification by the TRACE-like front-end), setting up an activation gradient. Second, similar recognition processes at the TRACE word level result in activation of the appropriate chunk layer node in the present model. Such recognition is assumed to occur by the time the second phoneme has reached the phoneme layer. This is feasible if it is further assumed that there is some sluggishness or time delay in arrival of phonemic input at the phoneme layer, and this is consistent with human data indicating that word recognition occurs well before all the acoustic input has arrived (e.g., Tyler, 1992). The result is that the correct chunk layer node becomes active while the node representing the first phoneme is still the most active node at the phoneme layer. Third, weights from the activated chunk node to the various constituent phonemes are strengthened automatically via Hebbian learning² during the process of cycling through the syllable template. This weight adjustment is an automatic process occurring even if the chunk is a known one. Fourth, adjustment of weights from the phonological store to the active chunk layer node also occurs automatically, even though the word is a familiar one.

Finally, it is assumed that the TRACE-like speech recognition processes operate not only on auditory input, but also on the outputs of speech planning (which, as noted above, feed back via an efference copy).

The broad structure of the model having been outlined, we now discuss how the model addresses the range of target phenomena set out previously as desiderata.

² There are well-known constraints on the kinds of learning that can be achieved with Hebbian weight adjustment (e.g., see McClelland & Rumelhart, 1988). We have chosen to use this simple learning procedure primarily because it provides a basis for one-shot learning, in contrast with more powerful gradient descent algorithms (such as back propagation, for instance), which usually require many training cycles to learn a set of mappings.

Processing and Learning Words

Word Learning

Word learning was first on our list of target phenomena. We now explain how the model accounts for the word learning data previously discussed.

Consider the case of the novel three-syllable word *wup-gaam-sif* being presented to the system; its pronunciation, in International Phonetic Alphabet notation, is [wUp-gam-slf]. We ignore suprasegmental information associated with the word form, such as its stress contour.

Word recognition processes operating on the string *wup-gaam-sif* will fail to activate any node at the chunk layer, and so a new chunk node is activated. Once it has been activated, it emits its context signal, and Hebbian learning occurs for each phoneme in the word. This learning adjusts connection weights between the chunk node's context signal and (i) the phoneme layer and (ii) the syllable template. Automatic Hebbian adjustment of weights between the semantics and chunk layers also occurs, as well as weights from the phonological store to the chunk layer.

As each phoneme is input at the phoneme layer, it is parsed into syllable structure by means of the syllable template.³ When the syllable template has been cycled through once, one syllable has been processed. The chunk node that was already active will now have weights to the phoneme layer representing the sequence *wupg*. A second cycle through the syllable template will result in the chunk node having weights to the sequence *wupgaams* at the phoneme layer. One more cycle through the syllable template will result in the chunk node having weights to the sequence *wupgaamsif* at the phoneme layer. At this point there is no further input to the phoneme layer. As a result, input to the phonological chunk layer stabilizes. There now exists a chunk node representing the newly learned word *wupgaamsif*. However, these weights are subject to decay. For the word to be learned permanently, weights must reach saturation. It is possible to offset the effects of weight decay through re-presentation of the word: every time the word is encountered, its weights are (automatically) strengthened. Re-presentation can be achieved either through actually encountering the word again in the environment or by rehearsal.

It should be noted that in the present model, words are simply nonwords whose weights have become strengthened. In all other respects, words and

³ As in the Hartley and Houghton (1996) model, the manner in which the syllable template parses input and output into syllable structure does not correspond with the linguistic Maximal Onset Principle, according to which *wupgaamsif* should be parsed as *wup-gaam-sif*. In fact, the syllable template leads to this being parsed as *wupg-aams-if*. This is a limitation of the syllable parsing mechanism, as Hartley and Houghton (1996) acknowledge. As they point out, the syllable template should be viewed as an approximation whose main purpose is to impose syllable structure and not to provide a detailed model of syllabification. For present purposes, it was not considered crucial to elaborate on these mechanisms.

nonwords are treated in the same way and are both represented at the phonological chunk layer. This contrasts with the suggestion made by Hartley and Houghton (1996) that words and nonwords might be represented at different layers.

Word Form Repetition

The second set of phenomena to be accounted for related to the repetition of word forms.

Following registration of a word form in the system, irrespective of whether it is a familiar or unfamiliar word, there will be an active node at the chunk layer. Note that weight adjustment (which is automatic, in the present model) will have occurred on presentation of the word form. To repeat the word form, the phonological store avalanche node is activated. This reactivates the appropriate chunk layer node, and this in turn sets up activation gradients at the phoneme layer. This enables articulation and hence repetition of the word form. Such repetition should be possible for several seconds after presentation of the word form, i.e., for as long as the various weights have not decayed too much. This constitutes the basic process whereby words and nonwords are repeated.

The effect of nonword length on repetition performance also has a natural explanation in the model. The greater the number of syllables in a nonword, the greater will be the decay of weights between the phonological store and the chunk layer node allocated to that nonword, and also the greater the decay of weights from the chunk layer to the phoneme layer. Consequently, longer word forms will be more poorly recalled (i.e., repeated).

Note that the phonological units (in the present model, phonemes) constituting a novel word form are all bound to a single phonological chunk layer unit. This is in contrast to the Hartley and Houghton (1996) model, in which there is no distinction between a sequence of phonological units (syllables, in that model) that constitutes a single novel word form and a sequence that represents a list of novel monosyllabic word forms.

Intentional Production of a Known Word

The third set of phenomena to be accounted for relates to the intentional production of known words.

The process of producing a word requires the activation of the appropriate chunk layer node. It is assumed that specific chunk nodes can be activated, via semantics. When the chunk layer node is activated, it will recreate its learned activation gradient over the phoneme layer, leading to sequential activation of the phonemes comprising the word. These activations will produce an output sequence of individual phonemes via operation of the competitive filter associated with the phoneme layer. These individual phoneme activations from the phoneme layer are assumed to form the input to expres-

sive programming and then to eventual production of the word. Overall, these processes constitute the model's account of the production of known word forms.

Verbal Short-Term Memory

In addition to the various aspects of word learning and processing discussed above, the model is designed to provide an account of processing in verbal short-term memory and of immediate serial recall in particular.

Immediate Serial Recall without Rehearsal

In the model, serial recall ability for lists of up to about five items is automatic and nonstrategic. During presentation of the sequence of list items, the sequence of activations caused at the chunk layer is automatically encoded via Hebbian adjustment of the weights from the phonological store avalanche to the chunk layer; at recall, reactivation of the phonological store avalanche leads to replaying of the sequence of chunk layer activations. This in turn leads to replaying of a sequence of phoneme layer activations. This is the ability involved when subjects report "reading out" the recall sequence from memory, as they commonly do for lists of up to five digits. Such memory involves no rehearsal. It corresponds to the initial part of the serial recall curve, where performance is almost perfect for up to approximately five digits (Guildford & Dallenbach, 1925). Thus in the model, "pure" serial recall ability is attributed to the phonological store.

The existence of verbal short-term memory abilities without rehearsal is supported by neuropsychological evidence. For example, despite neurological impairment precluding any rehearsal abilities, Howard and Franklin's (1990) patient MK had an auditory list memory capacity. This memory span was non-time-limited, but fell off quite sharply with more than three or four items. In the present model, such abilities are attributable to the phonological store.

The serial recall abilities just described in the model rely on weights from the phonological store avalanche to the chunk layer. However, they are also dependent on weights from the chunk layer to the phoneme layer and on the weights between the chunk layer and semantics. Word learning is also, of course, dependent on these latter two sets of weights. The common reliance of both serial recall ability and word learning on these weights thus constitutes one aspect of the relationship between the two abilities: they both depend on basic phonological processing.

Note that the present model incorporates two levels of sequencing: sequencing of the word forms constituting a list and sequencing of the phonemes within each word form. None of the earlier computational models have addressed sequencing at both of these levels. The Burgess and Hitch (1992) model addresses sequencing of the word forms within a list. However,

it does not represent the fact that word forms are temporally extended, and therefore does not address the sequencing of phonological units within each word form. The Hartley and Houghton (1996) model addresses the sequencing of phonological units (phonemes) within monosyllabic nonwords. However, it does not have a representation of familiar word forms, and so it does not deal with the sequencing of lists of known words.

Immediate Serial Recall with Rehearsal

For lists longer than about five items in length, serial recall is no longer virtually perfect, and strategic factors play an increasingly important role as list length increases. Human subjects exhibit two primary strategies: cumulative rehearsal and chunked rehearsal. In cumulative rehearsal, the subject attempts to rehearse the entire list, updating it as each digit is presented. In chunked rehearsal, the subject breaks the list into chunks, or groups of digits (typically, groups of two, three, or four digits); each of these chunks is rehearsed separately. These strategies most frequently rely on subvocal rehearsal, but subjects sometimes report using visualization of arrays of digits as an alternative (or additional) means of rehearsal. It is usually not possible to recall much more than five digits without use of some such strategy.

In the proposed model, such chunking processes involve the binding of a sequence of phonemes (constituting the recall stimuli) to a chunk-level node. Of course, this is the same process as in the model's account of word-learning, and this constitutes another important aspect of the relationship between the domains of vocabulary acquisition and verbal short-term memory. Below, cumulative rehearsal and chunked rehearsal are described in terms of the model, with respect to ISR of the list {8 3 9 5 1 6 7 4 2}.

Chunked Rehearsal

In chunked rehearsal, the subject waits until some number of digits have been presented and then repeats them as a chunk. Assuming a chunk size of three, processing in the model would be as follows. The phonemes constituting the digit 8 (the first digit in the list) enter at the phoneme layer, and word recognition processes lead to activation of the appropriate node at the chunk layer. Automatic adjustment of weights from the phonological store to the chunk layer occurs. This weight adjustment continues while the next two digits (3 and 9) enter the system. At this point, the phonological store has an encoding of this sequence {8 3 9}.

At this point, when a chunk has been presented, the digit sequence is "replayed" from the phonological store, just as if the list {8 3 9} were being recalled; this leads to sequential activation of the phoneme layer units constituting the sequence. Additionally, a new node is allocated at the chunk layer, and weights from it to the phonemes constituting the sequence {8 3 9} are strengthened. Thus a new chunk encoding this subsequence has been

created. Weights between the semantics and chunk layers are adjusted, creating a contextual semantics for this new chunk.

The sequence of phoneme layer activations created by recall in turn activates the speech planning system. This expressive activity can be recycled. Via word recognition processes, it reactivates the phonemes in the chunk {8 3 9}, as well as the chunk layer node allocated to this chunk. This constitutes rehearsal of the chunk. Each time rehearsal occurs, weights between the chunk node and the semantics layer will be restrengthened; this is the function of rehearsal in the model. This process can be repeated; each time, the chunk will be better learned, exactly as in the model's account of new-word learning.

While rehearsal of the first chunk occurs, other digits have been entering the system. These digits are encoded automatically by the phonological store. As this is an automatic processes, rehearsal of the first chunk can continue through this presentation. When three new digits have been presented ({5 1 6}), rehearsal of the first chunk {8 3 9} stops, and the strategic processes described above for the first chunk now begin for the second group of three digits: playback from the phonological store, chunking, and rehearsal.

Meanwhile, the first chunk needs to have been stored somewhere. In the model, each time rehearsal is initiated on a new chunk, the current state of the chunk layer is entered as a vector into the context maintenance queue associated with the chunk layer (see Fig. 5). By encoding these vectors, the queue encodes the sequence of chunks, as the most active chunk node at each point is the one that was being created or rehearsed. This queue comprises a pool of units, one of which is allocated whenever a new chunk starts being rehearsed. Thus each chunk is encoded by one of these queue units. Serial order among these units, and hence among the stored chunks, is encoded by the mechanisms described by Bairaktaris and Stenning (1992).

In this way, two chunks will be created for {8 3 9} and {5 1 6} and will be entered into the queue. While the last three digits of the list are being presented ({7 4 2}), the first two chunks are rehearsed, by means of removal from the queue, repetition, and reentry into the queue. Removal from the queue is a process whereby the vectors encoded in the queue are re-presented at the chunk layer. Note that vectors are removed from the queue in the same order that they were entered into the queue. Each vector recreates its pattern of activation over the chunk layer, thus enabling repetition of the most active chunk, viz., the one that was created. Repetition (rehearsal) of the chunk also leads to reentry of the chunk layer vector into the queue.

When list presentation is over, the two chunks {8 3 9} and {5 1 6} are again extracted from the queue and articulated. The last three digits are read out of the phonological store at the end, without any chunking or rehearsal.

Cumulative Rehearsal

The model's account of cumulative rehearsal differs from chunked rehearsal in the following way. Instead of chunking the list into groups of

digits, a strategy of cumulative rehearsal attempts to create a chunk encompassing all the list elements. At the beginning of presentation of a list such as {8 3 9 5 1 6 7 4 2}, a chunk layer node is allocated. Weights from it to the active phoneme nodes for the first digit (8) are strengthened, and the chunk is articulated. This leads to reactivation of the phoneme nodes for 8, and weight readjustment. Meanwhile, the next digit (3) is presented, so the chunk layer node now chunks the group {8 3}. The group is repeated, leading to weight adjustment. Meanwhile, the next digit comes in (9) and is added to the chunk. The chunk (now {8 3 9}) is repeated, and meanwhile the fourth digit, 5, comes in and is added to the chunk; the chunk is repeated. This procedure continues for each additional digit. Where cumulative rehearsal differs from chunked rehearsal is that further digits are also added to the same chunk. That is, the same chunk layer node is used as the chunk to which all the recall stimuli are sought to be bound. This is difficult because this process requires repetition of increasingly long digit strings, which tends to interfere with the newly incoming digits, making it difficult to add them to the chunk. This strategy is usually not very effective for more than about 7 digits.

Immediate Serial Recall Data

We now turn to accounting for the specific phenomena we targeted in an earlier section. The fourth set of phenomena related to immediate serial recall of lists of words.

Limits on span, the serial position curve, and chunking. In the model, limits on serial recall arise from various sources. Performance will be virtually perfect for lists of up to about five digits. Beyond this, constraints on performance will differ depending on what strategy is in use. Under cumulative rehearsal, limits on serial recall will arise from limits on the number of items that can be bound to a single chunk node (because of the decay of weights) as well as from limitations on articulation rate; consequently, ISR performance will exhibit a steady decrease from list length five upward. Under chunked rehearsal, capacity limitations will arise from articulation rate together with constraints on the number of vectors that can be placed in the queue. This is in line with Zhang and Simon's (1985) finding that span depends on the number of chunks as well as on articulation rate.

The model also provides an account of the shape of the serial position curve. Under cumulative rehearsal, primacy will occur because the context signal for a chunk is most differentiated at the beginning and end (Burgess & Hitch, 1992; Houghton, 1990). This same effect promotes recency effects, which are also supported by readout from the phonological store for the last one or two items. Under chunked rehearsal, the primacy effect arises from the effects of context differentiation (favoring the first position within a chunk) together with the fact that the first chunk will have been the one most rehearsed. The recency effect arises from a combination of the effects of con-

text differentiation (favoring the last position within a chunk) and phonological store support for the last chunk as a whole.

“*Articulatory loop*” effects. The model provides an account of the various specific phenomena that have inspired the articulatory loop framework.

The phonological similarity effect will arise because of the topographic organization of the chunk layer. During presentation of each item in a list, vectors of chunk layer activations are encoded in the phonological store, at each time step. If the list items are phonologically similar (e.g., *bat, cat, sat*), then the vectors to be encoded will be very similar at different time steps. The chunk layer vectors produced at different time steps during recall will therefore be very similar, increasing the likelihood that an incorrect chunk layer node may be activated.

The irrelevant speech effect will arise because it results in activation of extraneous nodes at the phoneme layer, and, if the irrelevant speech stimulus is a word, then also at the chunk layer. It is assumed that this irrelevant speech can be attentionally filtered so that activation of these nodes is relatively low. Nevertheless, the effect of the irrelevant speech is to increase the number of competing nodes at these various layers, thereby affecting (reducing) activation of the nodes representing actual recall stimuli, thus causing weights to be weaker than they would be in the absence of the irrelevant speech.

The word length effect arises with cumulative rehearsal because repetition of the whole list in between presentation of recall stimuli is even more difficult with long words than with short. With chunked rehearsal, word length will affect the number of rehearsals of a chunk that are possible before the next chunk must be rehearsed. For example, for the short-word list {*whale, harp, smile, knife, plant, crown*}, the chunk *whale-harp-smile* might be rehearsed three or four times while the remaining list items are being presented. For the long-word list {*crocodile, orchestra, hospital, telephone, photograph, tangerine*}, however, rehearsal of the chunk *crocodile-orchestra-hospital* might only be possible once or twice while the rest of the list is presented. Chunks in the short-word case will therefore have stronger weights to the syllable layer than in the long-word case and so will be better recalled.

Concurrent articulation will lead to reduced availability of expressive programming mechanisms and, if the concurrent articulation is overt rather than silent, to an irrelevant speech effect as well (Gupta & MacWhinney, 1995). In terms of the model, the reduced availability of expressive programming mechanisms reduces the number of times that chunks can be rehearsed, while the accompanying irrelevant speech effect lowers the activations of chunk nodes because of competition from the irrelevant speech stimuli. For these reasons, concurrent articulation will reduce memory span.

Under concurrent articulation, phonological similarity will have an impact in the same way as described above, so there will still be a phonological

similarity effect. Under concurrent articulation, external irrelevant speech will cause further degradation in ISR performance. Not only will chunks be less frequently rehearsed because of concurrent articulation, but they will also have lower activations because of competition from two kinds of irrelevant speech stimuli: those engendered by the concurrent articulations themselves (Gupta & MacWhinney, 1995) and the actual externally generated irrelevant speech stimuli (Hanley & Broadbent, 1987). Overall, chunks will be recalled even more poorly than under concurrent articulation alone.

Under concurrent articulation, reduction of the word length effect will occur because chunks can only be refreshed between concurrent articulations. The effect of this is to fix the refresh rate of chunks at a constant rate, irrespective of whether the words in them are long or short, so that word length effects will be reduced.⁴ To see this, assume that each concurrent articulation takes X msec and is repeated at the rate of once every second. Then in every second, there is a fixed $1000 - X = Y$ msec period during which to perform output processing for refresh of a recall stimulus chunk. It seems realistic to assume that Y is only long enough for expressive processing of one chunk, so that, irrespective of the length of the words in it, only one chunk can be rehearsed per time period. In case short words are short enough that a chunk can be rehearsed twice in time period Y , or in case long words are long enough that a chunk can be only partially rehearsed in time period Y , then there could still be word length effects, and so the account can accommodate the possibility mentioned in Baddeley (1986) of some remaining word length effect even under concurrent articulation. The magnitude, however, will be reduced.

Serial Recall for Words vs Nonwords

The fifth set of phenomena we targeted related to differences in the immediate serial recall of lists of words and nonwords.

For a nonword (i.e., a novel word form), no weights exist from a chunk node to the phoneme layer or to the semantics layer; these are what develop in learning the nonword. Moreover, weights from the allocated chunk node to the phoneme layer undergo decay. In immediate serial recall of nonwords, a sequence of nonwords is presented. Each new word form must be rehearsed a few times as soon as it is presented or the newly developed weights will decay too much. This need for immediate rehearsal for each new list item makes it difficult to perform chunking—rehearsal is needed to encode the most recent list item and is therefore less available for rehearsal of chunks. Even if a chunk can be created for a group of nonwords, it will be rehearsed less frequently, and its weights will therefore tend to decay more than those

⁴ This explanation was first proposed in Gupta and MacWhinney (1993); a similar explanation has been put forth by H. B. Richman and H. A. Simon (manuscript in preparation).

for a chunk of words would. For these reasons, memory span for nonwords will be inferior to that for words.

Another difference between serial recall of lists of words and nonwords is that, for word lists, the predominant errors are item errors, in which items in the list are transposed or wrongly ordered without transposition. For lists of nonwords, however, the predominant errors are interchanges of parts of word forms. The present model provides an account of why transposition errors would occur in lists of nonwords rather than lists of words. The reason is that chunk layer activations will be more stable for words than for nonwords, because there is stronger activation from semantics. During recall of a list of known words, therefore, when a particular chunk layer unit wins the competition at the chunk layer, it is likely that it will continue to win the competition at each time step until the full word has been produced. It is unlikely that it will be preempted by some other word form, which would result in substitution of phonemes from that incorrect word. In the case of nonword lists, the lesser stability of chunk layer activations increases the probability of such preemption, and therefore of transposition errors. By virtue of incorporation of the Hartley and Houghton syllable template, the present model can also account for why such transposition errors adhere to the constraints of syllable structure.

Thus the model provides an account of the superior ISR performance for lists of words than for lists of nonwords, and of the differences in error types.

NEURAL SUBSTRATES

We began this paper by pointing out that behavioral data, together with the theoretical framework of the working memory model, suggest a very simple schematic of the system underlying performance in both verbal short-term memory and vocabulary acquisition. In immediate serial recall, this system sets up a loop of activations (rehearsal) which can maintain phonological representations of the recall stimuli in an active state. In vocabulary acquisition, likewise, a similar rehearsal loop aids the formation of a phonological representation for a new vocabulary item.

We also pointed out that surprisingly little is known about what mechanisms might underlie such a processing system. In the previous section, we presented a computational model specifying what the nature of these mechanisms might be. In terms of this model, the relationship between word learning and immediate serial recall arises from two factors. First is the common reliance of these two abilities on the Chunk \rightarrow Phoneme weights and the Semantics \leftrightarrow Chunk weights. In addition, the two abilities are related in their common reliance on articulatory rehearsal mechanisms.

Given this hypothesized processing system composed of speech planning mechanisms and phonological representations, the obvious question is whether and how this system is related to the mechanisms of language pro-

cessing. It is important to ask to what extent the cognitive system we have hypothesized to be commonly implicated in immediate serial recall and vocabulary acquisition maps onto the fundamental mechanisms that perform speech processing in language.

There appears to have been little previous examination of this question. Although studies of immediate serial recall have long hypothesized a partnership between expressive mechanisms and phonological processes, there has been, as Monsell (1987) has noted, “. . . a curious reluctance to specify with any precision . . . [its] relation to ideas about the lexicon . . .”. Thus the behavioral data suggest the existence of a processing system composed of speech output and phonological representational/storage components, but say nothing about its relation to any other cognitive apparatus.

Here, too, the computational model we have proposed provides a clear statement of the relationship between this cognitive system and mechanisms of language processing more generally. In our view, immediate serial recall involves the opportunistic deployment of core language processing circuitry. A similar suggestion has previously been made by Howard and Franklin (1993); however, our model provides the first precise specification of such a relationship.

We can further ground this functional specification by examining the extent to which this cognitive system maps onto what is known about the neurophysiology of speech processing in language. Once again, the issue appears to have received little attention. Even in a recent review of the working memory model for a cognitive neuroscience audience (Baddeley, 1992), there is little discussion of the possible physiological bases of the articulatory loop or of its relationship to language processing [but see Shallice & Vallar (1990)].

This section attempts to identify candidate neural structures for the processing subsystems hypothesized in the functional account. Our strategy is, first, to review what is known of the neural substrates of speech processing. We begin by considering the likely neural substrates of articulatory programming in language. Next, we examine the likely neural substrate of the system of phonological representations. In the third section, we review other findings that seem difficult to reconcile with the data reviewed in the first two sections. There is currently no accepted theoretical framework that integrates these various data. In the fourth section, we suggest a conceptual framework within which these various findings can be accommodated. This concludes our review of the neural bases of speech processing. We then review neuropsychological evidence pertaining to immediate serial recall and vocabulary acquisition, examining whether the processing system we have specified at a computational level maps onto the neural circuitry we have specified for speech processing.

Some clarifications are in order before we begin. First, the classical view regarding the neural bases of language has centered around the notions of Broca's and Wernicke's aphasia, which seemed to indicate distinct types of

aphasia arising from lesions at different cortical sites (Mackay, Allport, Prinz, & Scheerer, 1987). Anterior areas, especially Broca's area, were viewed as responsible for language production, and posterior areas, especially Wernicke's area, were responsible for language comprehension. This view was incorporated into the influential Geschwind model and continues to be adopted in current treatments (e.g., Mayeux & Kandel, 1985). In recent years, however, a variety of findings have challenged this simple dichotomy (Mackay et al., 1987). Damage confined to Broca's area does not cause Broca's aphasia, which is caused, rather, by damage encompassing both the classical anterior and posterior language areas (Mohr, 1976). Electrical stimulation work has demonstrated that stimulation at a single site can interfere with both the perception and production of speech (Ojemann, 1983). Both Broca's aphasics and Wernicke's aphasics have been shown to exhibit both production and comprehension deficits (Blumstein, 1990). Thus the classical mapping of a production-comprehension dichotomy onto an anterior-posterior dichotomy does not hold up (Blumstein, 1995).

We will therefore steer clear of issues about language comprehension vs. language production and confine ourselves to considering the neural bases of the speech processing of single spoken words. In considering such processing, two further distinctions need to be made. The first is a distinction between input and output speech processing. The second is a distinction between phonology and phonetics.

"Input" speech processing refers to the processing necessary for identifying a spoken word during comprehension and accessing its associated properties (such as meaning); "output" processing refers to the speech processing involved in the generation of a spoken word from its meaning (Monsell, 1987).

In linguistics, a fundamental distinction is made between "phonetics" and "phonology" (e.g., Crystal, 1991). Phonetics is the study of the range of sounds used in speech, across languages. Phonology is the study of the way sounds pattern within particular languages. Roughly, phonetic processes are concerned with the identification or production of various features such as voicing or frication, whereas phonological processes involve knowledge of the sound pattern of a particular language and use of this knowledge in the process of identifying or producing spoken words.

For example, in output speech processing, phonological processes lead to such phenomena as phoneme insertion, deletion, substitution, and assimilation. These phenomena occur in essentially all languages, but pattern differently in different languages. Normal speech errors are assumed to arise from the faulty operation of these phonological processes. In contrast, phonetic output processing is viewed as more closely related to the articulatory realization of speech output. As an example, the degree of aspiration with which the phone [t^h] is produced in saying *teams* would be a phenomenon of phonetic processing.

In input speech processing, phonological processes are involved, for example, in discriminating between the spoken words *pat* and *bat*; phonetic processes are involved in a phenomenon such as categorical perception, which involves the extraction of spectral patterns corresponding to the phonetic categories of speech (Blumstein, 1995).

We are now ready to begin consideration of the neural bases of the speech processing of single spoken words, both input and output, both phonetic and phonological. What are the data?

Frontal Articulatory Mechanisms in Language

As Keller (1987) has noted, the frontal lobe has undergone the greatest phylogenetic expansion in humans; it is motorically dominant, is suited to fine motor control, and is the site of a disproportionately large area concerned with vocal tract control. The primary and secondary motor cortices thus appear well equipped for mediation of fine motor control in the vocal tract and for speech processing. The primary motor cortex (M1) is the Brodmann area (BA) 4, constituting the precentral gyrus. The secondary motor cortex is less clearly circumscribed, but is usually considered to include BA 6, whose lateral and medial surfaces are respectively termed premotor cortex and the supplementary motor area (Kupfermann, 1985).

Broca's area (Brodmann area 44) lies immediately anterior to the premotor cortex and has a long history of being viewed as the secondary motor area specialized for speech motor control/programming (Geschwind, 1979; Liepmann, 1915; Mayeux & Kandel, 1985; Mohr, 1976). From careful and exhaustive examination of cases involving focal Broca's area lesions, Mohr (1976) concluded that the resultant deficit is an "apraxia of speaking," an inadequacy in skilled execution of movements, and that the role played by Broca's area is that of a premotor association cortex region concerned with movements required for articulation. Support for this conclusion comes from models of the neural correlates of speech programming and execution. For example, based on a review of neurophysiological and neuroanatomical evidence, Gracco and Abbs (1987) have proposed a model of the neural correlates of speech programming and execution processes in which the cortical area subserving early articulatory planning is lateral premotor cortex. An equivalent formulation is provided by Barlow and Farley (1989). It should be noted that Gracco and Abbs' (1987) view of lateral premotor cortex broadly includes Broca's area. This coincides with the conclusion reached by Mohr (1976), viz., that the role played by Broca's area is that of a premotor association cortex region concerned with acquired skilled oral, pharyngeal, and respiratory movements. Mohr (1976) has also noted that no cytoarchitectonic justification for assignment of a special number (BA 44) to Broca's area was ever published by Brodmann in a human case and that there is no evidence for a distinction between Broca's and other premotor areas.

This view of Broca's area needs to be carefully distinguished from two other views that have been prevalent in neuropsychology at various times. First, it contrasts with views that have equated Broca's area damage with Broca's aphasia. Mohr (1976) found that, contrary to the classical view, Broca's area is not responsible for Broca's aphasia, which is the product of much more extensive damage involving essentially the whole of perisylvian cortex in frontal, temporal, and parietal cortex: essentially, the region supplied by the middle cerebral artery. A number of studies have confirmed and extended Mohr's (1976) conclusions (Naeser & Hayward, 1978; Kertesz, Harlock, & Coates, 1979), and according to Mateer (1989), there is now general agreement that lesions confined to Broca's area result in only transient aphasia and not in the chronic condition of Broca's aphasia. Second, it contrasts with views, summarized, for example, in Kean (1985), that have sought to localize the functions of grammar and syntactic processing to Broca's area, but that are now largely discredited: see Bates, Thal, and Janowsky (1988) for review.

An articulatory planning role for secondary motor areas is also supported by a variety of other findings. Evoked potentials of electrophysiological activity have been shown to occur in Broca's area milliseconds before actual articulation begins (McAdam & Whitaker, 1971). Similarly, neural activity in the basal ganglia has been recorded approximately 100 msec before speech (Borden, 1980); the basal ganglia are known to project to premotor speech areas (e.g., Gracco & Abbs, 1987). Further evidence comes from positron emission tomographic imaging studies showing that actual and imagined right hand movement activates cortex near the left-side-buried Sylvian cortex (approximately, BA 43) (Fox, Burton, & Raichle, 1987). Of particular significance is the fact that imagined movements produced the same activations, strongly suggesting planning activity that stopped short of motor execution. This evidence makes it likely that premotor activations play a similar role in speech motor coding (Petersen et al., 1989).

The hypothesis we adopt, therefore, is that the cortical mechanisms implicated in articulatory programming, planning, and control are composed of Broca's area, the premotor cortex, and supplementary motor area (Gracco & Abbs, 1987). These articulatory mechanisms constitute part of the neural substrates that realize output phonetic processing in speech (Blumstein, 1995). Execution of speech, i.e., actual overt articulation, is controlled by primary motor cortex, whose pyramidal cells output to motoneurons that are mapped to muscles of the vocal tract (Barlow & Farley, 1989).

Phonological Representations

The neural correlates of input phonological processing have been located classically in temporal and parietal cortex, in and around Brodmann areas 22 (the classical Wernicke's area), 39 (angular gyrus), and 40 (supramarginal

gyrus) (e.g., Mayeux & Kandel, 1985). These localizations seem largely borne out by the literature on brain imaging using positron emission tomography (PET).

For example, in work reported by Zatorre et al. (1992), there was a subtractive hierarchy of conditions, as follows. In the baseline condition, subjects were passive, in silence. In the noise condition, subjects listened to pairs of noise bursts and pressed a key to alternate pairs. In the passive speech condition, subjects listened to pairs of syllables and pressed a key to alternate pairs. In the phonetic judgment condition, subjects listened to the same pairs of syllables as in the passive speech condition and pressed a key whenever both members of a pair ended in the same consonant (which was 50% of the time). In the pitch judgment condition, subjects listened to the same pairs of syllables as in the passive speech and phonetic judgment conditions and pressed a key whenever the second member of the pair had a higher pitch than the first (which was 50% of the time).

In the relevant subtractions, the phonological processing of speech sounds was found to activate temporal regions, while more complex phonological processing requiring phonetic judgments additionally activated the supramarginal gyrus (BA 40) and Broca's area.

Demonet et al. (1992) employed three experimental conditions involving auditory presentation of stimuli. In the baseline tones condition, subjects monitored triplets of tones for rising pitch and indicated detection by pressing a button with the right index finger. The other conditions employed the same target detection response. In the phonemes condition, subjects monitored nonwords for the occurrence of the phoneme /b/ in a word in which /d/ had occurred in a previous syllable. In the words task, subjects monitored adjective-noun pairs for occurrence of a noun denoting a small animal (smaller than a chicken or a cat), preceded by a "positive" adjective (e.g., *kind*, vs *horrible*).

Once again, the subtractions reveal that complex operations on phonological representations (as in the phonemes task) are accompanied by Broca's area activation. The angular/supramarginal gyrus activation (BA 39/40) appears for processing of phonological forms that are specifically lexical, as in the words task.

To summarize, PET studies have shown that perception of phonologically structured auditory stimuli activates nonprimary superior temporal areas bilaterally (BA 22/42, 39, 40) (Demonet et al., 1992; Howard et al., 1992; Petersen et al., 1989; Zatorre et al., 1992). Tasks requiring phonological judgments additionally activated Broca's area (left BA 44) (Demonet et al., 1992; Zatorre et al., 1992), while actual motor output produced activations of area 6 and areas 45, 46, and 47 (Petersen et al., 1989). These results not only support the view of posterior temporal/parietal cortex as specialized for input phonological processing, but also provide further evidence of the

role of anterior premotor cortices (BA 6, 44, 45, 46, 47) in articulatory programming and control.

Evidence for Anterior–Posterior Interactions

Our review so far seems consistent with a modified form of the classical view of brain and language. Specifically, it appears that even though language comprehension and production do not map onto the posterior and anterior language areas, perhaps input and output speech processing do. Broca's area, it appears, does output speech processing, and temporoparietal cortex does input speech processing. It also appears that phonetic judgments are specifically dependent on anterior language areas.

However, a variety of data challenge this simple picture as well (Blumstein, 1990, 1991, 1995). In particular, it has been shown that almost all aphasics, regardless of clinical type, exhibit difficulties in output phonology. Aphasics' difficulties in producing correct speech are regarded as difficulties in "phonological planning" (rather than "phonetic" processing difficulties) if the errors exhibit the same kinds of phonological processes as occur in normal speech. In aphasic speech, examples of such errors in output phonology include phoneme substitution (*teams* → *keams*), phoneme deletion (*green* → *geen*), phoneme insertion (*see* → *ste*, *apple* → *papple*), phoneme exchange (*degree* → *gedree*), progressive assimilation (*Crete* → *kreke*), and regressive assimilation (*Crete* → *trete*). Despite the regularity of such phonological errors, the particular occurrence of an error cannot be predicted. Importantly, this pattern of deficit in output phonology appears to be similar across types of aphasic lesion. However, Broca's aphasics and anterior aphasics do exhibit a greater degree of severity of such impairment than do posterior aphasics (Blumstein, 1995).

Nearly all aphasic patients also show difficulty in input phonology. Most studies exploring the speech perception abilities of aphasic patients have focussed on such patients' ability to perceive phonemic contrasts (e.g., discriminating between the auditory forms *pear* and *bear*), and nearly all aphasic patients exhibit some problems in such phonological discriminations (Blumstein, 1991, 1995). Other studies have found aphasics to be impaired in selecting the appropriate picture corresponding to an auditorily presented word, when the picture choices include pictures of items with phonologically similar names, for example, the auditorily presented word might be *pea* and the picture set might include pictures of a *pea*, *T*, and *key* (Blumstein, 1990). Aphasics show impairment on this task, suggesting reduced discriminability between input phonological representations. These discrimination or picture–choice abilities consistently appear to be impaired in aphasics, with no indication of systematic differences across type of aphasia or lesion site, or as a function of auditory language comprehension ability (Blumstein, 1991).

What these results suggest is that there is no simple mapping between input and output phonology, on the one hand, and anterior and posterior

language areas, on the other hand. Both input and output phonology are dependent on both anterior and posterior areas. Furthermore, electrical stimulation work has demonstrated that stimulation at a single site can interfere with both the perception and the production of speech (Ojemann, 1983).

There appears to be more evidence for anterior–posterior differentiation in phonetic impairments, however. In output phonetic processing, anterior damage, especially to Broca's area, appears to impair the ability to produce phonetic dimensions that require the timing of two independent articulators, i.e., speech–motor planning and coordination abilities (Blumstein, 1995). Patients who exhibit such difficulties also have impairments in timing relations between syllables; for example, they do not show a normal decrease in the duration of a root syllable as word length increases. They also show an inability to increase articulation rate beyond a certain point. This is particularly significant for the present discussion, because it provides evidence for the role of Broca's area in determining articulation rate and hence rehearsal rate. In contrast, output phonetic processing does not display such timing deficits under posterior damage. Posterior aphasics do, however, exhibit subtle subclinical phonetic impairments—for example, increased variability in the implementation of phonetic parameters such as vowel formant frequency. Overall, anterior damage produces substantially greater output phonetic impairment than does posterior damage (Blumstein, 1995). Posterior damage can, of course, lead to severe disruptions of output phonology, as in fluent jargonaphasia, even if there is no anterior cortical damage. However, in such cases there is almost always damage to subcortical structures, and so such deficits are not inconsistent with output phonology being more dependent on anterior than on posterior cortical areas.

It also appears that there are differences in the effects of anterior and posterior damage on input phonetic processing, for example, in phonetic categorization. In normal subjects, the lexical status of the environment in which a phoneme occurs affects its phonetic categorization. For example, in an acoustic continuum between the word forms *tash* and *dash*, varying only in voice–onset time of the first segment, the categorization boundary between /t/ and /d/ shifts so that there are more /d/ responses than usual. Similarly, in an acoustic continuum between the word forms *dask* and *task*, there are more /t/ responses than usual. Broca's aphasics show a larger-than-normal lexical effect, suggesting that they are relying on lexical status to a greater extent than normal subjects. Posterior aphasics, however, do not show a lexical effect, suggesting that lexical information does not affect their phonetic categorization (Blumstein, 1995).

These findings suggest that both input and output phonetic processing also depend on both anterior and posterior areas; however, there may be a greater degree of localization than in phonology, with input phonetic processing relatively more dependent on posterior areas and output phonetic processing relatively more dependent on anterior areas.

Neurophysiological data also suggest extensive interconnectivity between the anterior and posterior language areas. Traditionally, the arcuate fasciculus has been considered to transmit information from posterior to anterior language cortices (e.g., Mayeux & Kandel, 1985). According to Seldon (1985), little is actually known about the fasciculus, such as the number of axons it contains, or whether it is a one-way or reciprocal projection, but analogy with other corticocortical projections suggests that it contains bi-directional connections. Mesulam (1990) has suggested that regions corresponding to Broca's and Wernicke's area in the rhesus monkey may each be connected to the same set of at least eight interconnected cortical areas. According to Pandya and Yeterian (1985, p. 19), auditory association areas project to frontal, parietotemporal, and paralimbic association areas; in particular, the posterior-most region of the superior temporal gyrus projects to premotor frontal cortex and to parietotemporal cortex. Similarly, the posterior-most part of parietal association cortex projects to prefrontal as well as premotor cortex and to the superior temporal sulcus (Pandya & Yeterian, 1985, p. 28). Finally, premotor regions of frontal cortex project to rostral parietal cortex and posterior superior temporal cortex (Pandya & Yeterian, 1985, p. 40), completing the loop between temporoparietal and premotor cortices. This neuroanatomical evidence, together with the neuropsychological and neurophysiological evidence we have already cited, justifies treatment of the frontal premotor and temporoparietal regions as forming an interconnected system.

To summarize, these various data suggest that both input and output phonology and phonetics are dependent on both anterior and posterior language areas, and this in turn suggests that the speech processing system may be interactive in nature, rather than modularly divided into "input" and "output" phonology or phonetics. Nevertheless, the evidence reviewed in the previous two sections also indicates that there are specializations within this interconnected system. All of these findings can be reconciled by assuming that the speech processing system is interactive in nature, but that input and output phonetics are more loosely coupled than input and output phonology. In the next section, we elaborate this view.

Reconciling the Data

The data reviewed in the preceding three sections can all be accommodated by assuming that the speech processing system consists broadly of left perisylvian cortex and that various aspects of speech processing map onto this substrate in different ways, as follows.

1. Aspects of speech processing relating to articulatory implementation and planning ("output phonetics") map onto Broca's area.

Mohr's (1976) careful examination of cases involving focal Broca's area lesions concluded that the role played by Broca's area is that of a premotor association cortex region concerned with movements required for articula-

tion. Posterior aphasics who do not have Broca's area damage do not exhibit these same kinds of articulatory planning impairments (Blumstein, 1995). Under "Frontal Articulatory Mechanisms in Language," we also reviewed other evidence that Broca's area plays an articulatory planning role (e.g., Barlow & Farley, 1989; Gracco & Abbs, 1987; McAdam & Whitaker, 1971). Blumstein (1995) has also concluded that phonetic processing depends on specific neural structures, and on Broca's area in particular. Output phonetic processing is of course required for actual speech output, and this is confirmed by PET studies showing Broca's area activation during tasks requiring speech output (Petersen et al., 1989). Such activation has also been revealed during tasks requiring covert speech, both using PET (in a silent picture-naming task: H. Chertkow et al., manuscript in preparation), and using fMRI (in a task in which subjects had to silently generate words beginning with a given letter: Cuenod et al., 1993; Hinke, Hu, Stillman, & Ugurbil, 1993; Rueckert et al., 1993).

2. The output phonetic processing subserved by Broca's area can be invoked for operations on phonological representations (for example, phoneme judgments).

In a PET study by Zatorre et al. (1992), the phonological processing of speech sounds was found to activate temporal regions, while more complex phonological processing requiring phonetic judgments additionally activated the supramarginal gyrus (BA 40) and Broca's area. A PET study by Demonet et al. (1992) showed that when subjects monitored nonwords for the occurrence of the phoneme /b/ following the occurrence of /d/ in a previous syllable, Broca's area was activated. Other PET imaging work (Paulesu et al., 1993a, 1993b) found Broca's area activation in a rhyme judgment task; this is consistent with experimental evidence suggesting that articulatory processes are involved in making rhyme judgments (Besner, 1987).

Thus Broca's area has been found to be activated in tasks requiring phonological judgments (Demonet et al., 1992; Zatorre et al., 1992). On the other hand, recent PET studies have not found Broca's area activations in purely "receptive" phonological processing tasks in which subjects listened to phonological stimuli but did not have to make judgments about them (Demonet et al., 1992; Howard et al., 1992; Petersen et al., 1989; Zatorre et al., 1992), although an earlier PET study did report frontal activation during comprehension (Lassen, Ingvar, & Skinhoj, 1978; Lassen & Larsen, 1980).

What is common to all the tasks that did show Broca's area activation is perhaps that they require operations on phonological representations. For example, judging whether the phone /b/ occurred in a word following a /d/ requires at minimum segmentation of the word into phones; it probably also requires additional operations such as similarity matching. Apparently, output phonetic processing is required in order to perform such operations. One possibility is that making phonetic judgments requires articulatory planning of the sound which is to be detected—a matching of input and output.

However, whatever the manner of involvement, it seems clear that output phonetic processing can be involved in making phonemic judgements.

Earlier, we cited evidence that lexicality effects in phoneme identification are differentially impacted by anterior and posterior lesions (Blumstein, 1995). Broca's aphasics show a larger-than-normal lexical effect, suggesting that they are relying on lexical status to a greater extent than normal subjects. The present hypothesis helps explain this: if output phonetic processes (subserved by Broca's area) are necessary for segmentation operations, and therefore for phonetic judgments, then Broca's aphasics would tend to rely on other sources in making such judgments.

3. Aspects of speech processing relating to auditory phonetic analysis map onto primary auditory cortex ("input phonetics").

Primary auditory cortex, known as AI, or Heschl's gyrus (Brodmann area 41), is located on the upper face of the temporal lobe, largely in the Sylvian fissure. It has a tonotopic organization, with isofrequency contours running along the dorsal-ventral direction (Evans, 1982; Gulick, Gescheide, & Frisina, 1989; Harrison & Hunter-Duvar, 1989; Yost & Gourevitch, 1987). There is also a columnar organization, in which cells have similar ear dominance within a cortical column (Gulick et al., 1989). Although little is known about its role in speech perception, it seems safe to assume that phonetic processing depends in some way on AI.

4. Less peripheral aspects of speech processing map onto an interconnected system composed of other perisylvian areas of temporoparietal cortex and other perisylvian areas of anterior premotor cortex. This anterior-posterior system constitutes the neural substrate of phonological representation ("input phonology" and "output phonology").

In the previous section, we reviewed evidence that aphasic patients exhibit difficulties in output phonology regardless of clinical type, which indicates that output phonology is dependent on both the anterior and posterior language areas; we also reviewed evidence that aphasic patients show difficulty in input phonology irrespective of lesion site, suggesting that input phonology is also dependent on both the anterior and posterior language areas (Blumstein, 1990, 1991, 1995). We noted also that neurophysiological evidence suggests interconnectivity between the anterior and posterior language areas, consistent with an interactive view of input and output phonology (Mesulam, 1990; Pandya & Yeterian, 1985; Seldon, 1985).

5. Anterior areas appear to play a greater role than posterior areas in output phonology, while posterior areas appear to play a greater role than anterior areas in input phonology. The evidence for the first claim comes from the finding that impairments of output phonology are more severe in Broca's aphasics and anterior aphasics than in posterior aphasics (Blumstein, 1995).

In considering the second claim, recall that input phonology is thought to depend on both anterior and posterior areas because of the finding that both

anterior and posterior aphasics show difficulties in input phonology. Two points are worth noting, however. First, the tests of input phonology in fact required phonetic discrimination (perception of phonemic contrasts) and were therefore probably not tests of purely input phonology. As we have argued above, making such judgments may require the use of output phonetic processing mechanisms subserved by Broca's area, and it is therefore not surprising that performance on these tasks is affected by both anterior and posterior lesions. The second point to note is that, in purely "input" phonological processing tasks not requiring phonetic judgments, PET studies have found the activation of superior temporal areas, but not Broca's area (Demonet et al., 1992; Howard et al., 1992; Petersen et al., 1989; Zatorre et al., 1992). These considerations together suggest that input phonology may rely to a greater extent on posterior areas than on anterior areas.

The view of the speech processing system we have just outlined can be made more concrete by envisaging it as a set of maps. The following discussion incorporates ideas from Grossberg (1986), who has discussed networks in which sensory and motor potentialities are integrated into a unitary system, and is also similar to Miikkulainen's (1990) use of multiple topologically organized maps. The terminology we adopt is that of Grossberg (see "Learning Serial Order"). The conceptualization is shown in Fig. 6a.

- The system is organized into multiple maps, or fields. The notation $\mathfrak{S}_M^{(i)}$ denotes the i th motor field in a hierarchy, and $\mathfrak{S}_S^{(i)}$ denotes the i th sensory field in a hierarchy.

- The hierarchy shown on the left represents a set of motor maps for speech. The lowest level in the hierarchy is field $\mathfrak{S}_M^{(1)}$, which represents terminal motor maps of a motor act; such a map specifies the terminal lengths of target muscles. The next level in the hierarchy is $\mathfrak{S}_M^{(2)}$, which has unitized representations of patterns of activation over $\mathfrak{S}_M^{(1)}$; such representations develop by instar learning. Also, these unitized representations at $\mathfrak{S}_M^{(2)}$ have the property that they can recreate their activating patterns of activity at $\mathfrak{S}_M^{(1)}$; this ability arises via outstar learning. These instar and outstar abilities are depicted as bidirectional connectivity between $\mathfrak{S}_M^{(1)}$ and $\mathfrak{S}_M^{(2)}$.

Further levels $\mathfrak{S}_M^{(3)}$ through $\mathfrak{S}_M^{(n)}$ in the hierarchy represent further levels of chunking in output phonology, for example, phonemes, syllables, and words. Each pair of layers has bidirectional instar-outstar connectivity.

Field $\mathfrak{S}_M^{(1)}$ represents motor cortex and $\mathfrak{S}_M^{(2)}$ represents Broca's area. Fields $\mathfrak{S}_M^{(3)}$ through $\mathfrak{S}_M^{(n)}$ represent other maps in premotor and frontal cortices.

- The hierarchy shown on the right represents a series of sensory maps for audition. For example, the lowest level in the hierarchy is field $\mathfrak{S}_S^{(1)}$, which represents phonetic feature discriminations. $\mathfrak{S}_S^{(2)}$ represents a level of perceptual processing whose properties give rise to such phenomena such as categorical perception. Further levels $\mathfrak{S}_S^{(3)}$ through $\mathfrak{S}_S^{(n)}$ in the hierarchy represent further levels of chunking in input phonology, for example, phonemes, syllables,

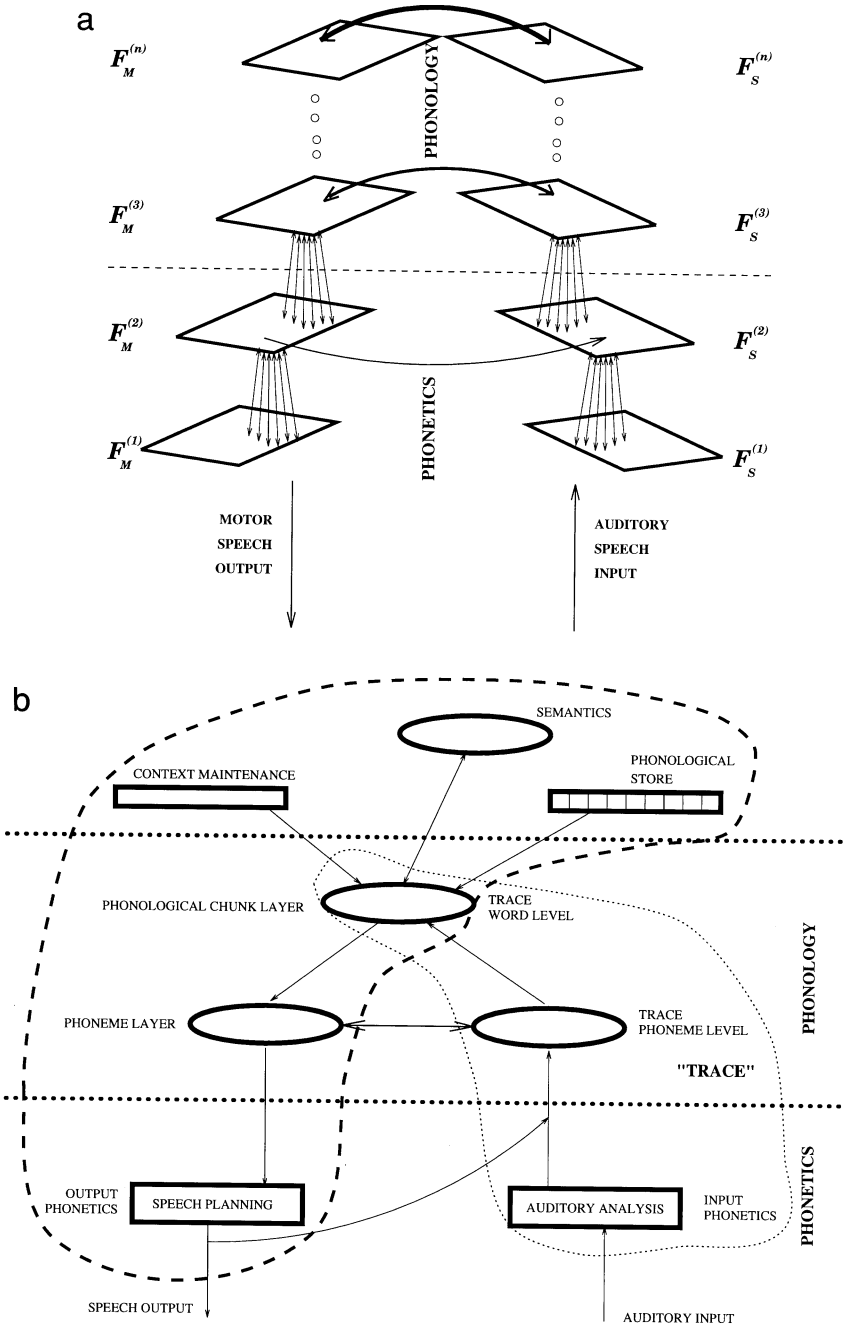


FIG. 6. Situating the model. (a) An interactive view of speech processing and input and output phonology. (b) Mapping the model onto the interactive view.

bles, and words. Each pair of layers has bidirectional instar–outstar connectivity, as described above.

Field $\mathfrak{S}_s^{(1)}$ represents primary auditory cortex, field $\mathfrak{S}_s^{(2)}$ represents secondary auditory cortex, and fields $\mathfrak{S}_s^{(3)}$ through $\mathfrak{S}_s^{(n)}$ represent other association maps in temporo-parietal cortex.

- At any level, bidirectional Hebbian learning occurs between the sensory and motor maps. Thus patterns of activation that occur together at the two maps can subsequently recreate each other. Note that such learning is constrained by the patterns of connectivity that actually exist between the sensory and motor maps at a particular level.

- The distinction between phonetic and phonological processing is interpreted in terms of the level of neural map. The distinction is drawn between fields $\mathfrak{S}_M^{(1)}$ and $\mathfrak{S}_M^{(2)}$, and $\mathfrak{S}_M^{(3)}$ through $\mathfrak{S}_M^{(n)}$, for output processing; and between $\mathfrak{S}_s^{(1)}$ and $\mathfrak{S}_s^{(2)}$, and $\mathfrak{S}_s^{(3)}$ through $\mathfrak{S}_s^{(n)}$, for input processing. This conceptualization emphasizes the fact that the distinction between phonetics and phonology is an analytically imposed one, based on level of representation, rather than one that reflects real differences in the nature of processing.

- At each level of the hierarchy, there is a certain degree of interconnectivity between input and output maps. The greater the interconnectivity, the greater the degree of coupling. At the phonetic levels, there is little interconnectivity and coupling. The degree of connectivity is shown as increasing at higher levels in the hierarchy.

This helps reconcile the data indicating that higher-level phonological processing is impaired by both anterior and posterior damage with the evidence suggesting that phonetics and “articulatory planning” can be identified with premotor cortex and Broca’s area: lower levels of the system are less interactive than higher levels.

- Broca’s area is represented as one level on the output side of the system. We have already reviewed evidence that Broca’s area is the neural substrate of articulatory planning, also identifying this level with output phonetics (Blumstein, 1995). The consequence of activation of this level of the output system is that internal feedback is delivered to $\mathfrak{S}_s^{(2)}$. This forms the basis for the “refresh” of phonological representations.

The mapping of our model onto this hierarchy of maps should be clear from comparison of Fig. 6a and 6b. It should be clear that the Phoneme Layer in Fig. 5 and in Fig. 6b is equivalent to field $\mathfrak{S}_M^{(3)}$ in Fig. 6a. However, the computational model we have described here abstracts away from interactions between sensory and motor maps at the phoneme level; it also assumes that sensory and motor maps are completely merged at higher phonological levels of the hierarchy such as the word form level (i.e., the phonological chunk layer). Blumstein’s (1995) conclusion that there is no distinction between input and output phonological representations amounts to the same assumption: that at some level of the system, the coupling is so strong as to form a single system.

Mapping the Model onto the Brain

Above, we have outlined our view of the neural bases of speech processing in language, showing how it reconciles the available data. We suggest that these same substrates constitute the speech planning and phonological representation components of the cognitive system we described in presenting our unified computational model. The evidence supporting this claim comes from deficits of digit span and vocabulary acquisition in patients with various kinds of neurological damage, which suggests that these processing abilities are impacted in different ways by different lesions (Baddeley et al., 1988; Shallice & Vallar, 1990; Waters et al., 1992) as well as from neuroimaging studies that have examined verbal short-term memory (Awh et al., 1994; Paulesu et al., 1993a, b; Petrides et al., 1993).

These data suggest the mapping of elements of our cognitive model onto the neural substrates of speech processing shown in Fig. 7, in which the correspondence between functional elements of the model and their hypothesized neural substrates is denoted by shading. The operation of this system is hypothesized to be as follows. Automatic serial recall abilities are subserved by regions of posterior temporal/inferior parietal cortex, which enable sequencing of phonological representations that have been activated in perisylvian association cortex. During rehearsal, frontal articulatory planning mechanisms are additionally employed, to refresh these phonological representations. If grouping strategies are employed, dorsolateral prefrontal areas may additionally be involved. In vocabulary acquisition, the imitative rehearsal strategies observed in children (Studdert-Kennedy, 1986; Kuhl & Meltzoff, 1982; Kuhl, 1991) serve the purpose of refreshing temporary phonological system activations, leading eventually to formation of a long-term phonological representation.

We now examine each part of this mapping in more detail.

Phonological Representation and Storage

In the model, the phonological representation system, composed of the Chunk, Phoneme, and Semantics Layers, is a fundamental component underlying verbal short-term memory and word learning. The STM avalanche is also a crucial component of the present model and corresponds to the working memory model's phonological store. These components map onto an interactive neural system that is importantly dependent on temporo-parietal cortex but that also encompasses anterior perisylvian regions.

Patients with posterior damage in general appear to suffer from span deficits (Risse et al., 1984). Furthermore, in reviewing the neuropsychological syndrome of "pure STM" deficit, which involves reduced auditory-verbal short-term memory in the absence of other major language and cognitive deficits, Shallice and Vallar (1990) conclude that the condition reflects impairment to a short-term "input phonological store" and, based on clinical-

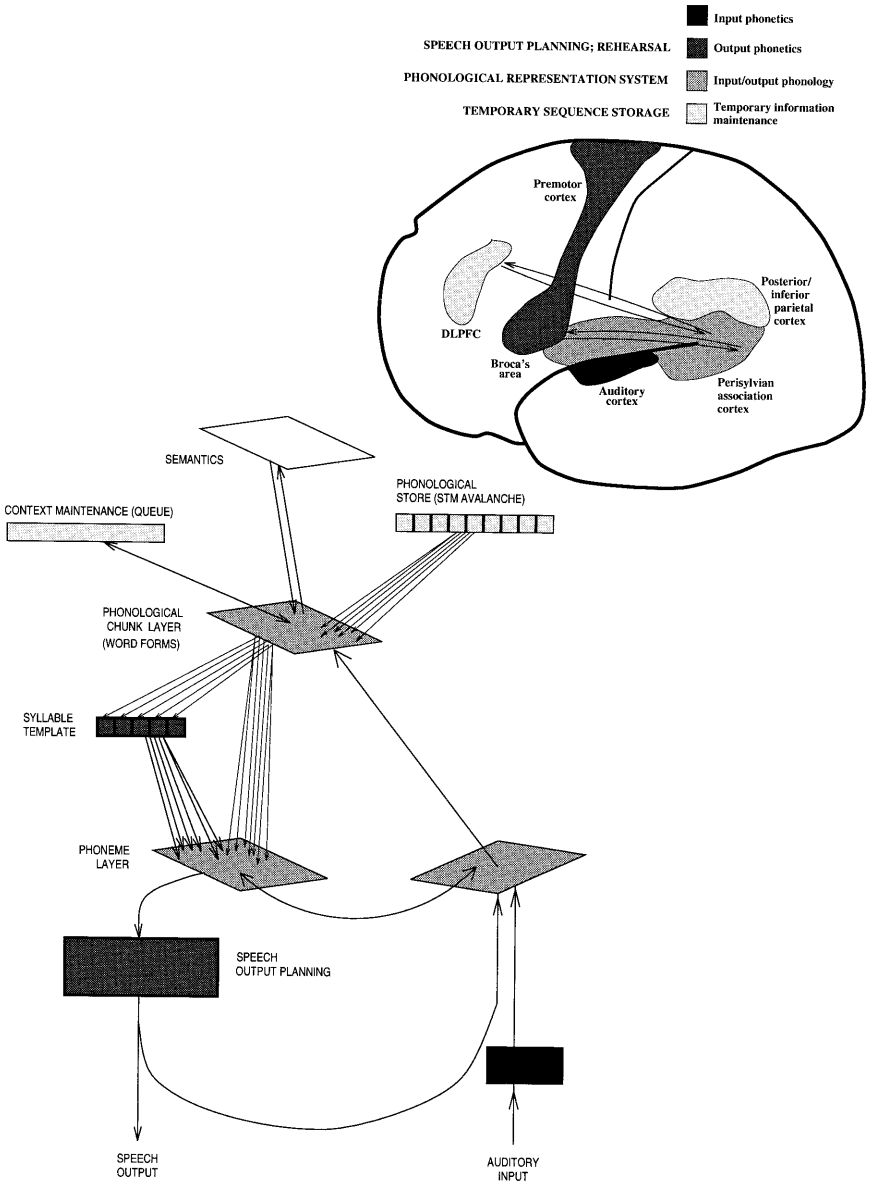


FIG. 7. Mapping of the proposed functional system onto frontal–temporoparietal neural substrates.

anatomical correlations, that the anatomical region compromised in this deficit is left inferior parietal cortex (angular gyrus and supramarginal gyrus). In patients with preserved span abilities, these areas seem quite consistently spared (Shallice & Vallar, 1990). This suggests that subsets of left temporoparietal cortex (e.g., left inferior parietal cortex) may be particularly specialized for the temporary storage of phonological information and thus are particularly crucial for verbal short-term memory. Supporting evidence that areas of temporo-parietal cortex play a role in the temporary maintenance of information comes from single-cell recordings in primates, which showed memory-related planning activity in posterior parietal cortex (Gnadt & Anderson, 1988), and from a PET study involving a verbal short-term memory task which revealed a supramarginal focus of activation, which the authors interpreted as the locus of phonological storage involved in the verbal short-term memory task (Paulesu et al., 1993b). What about vocabulary acquisition under damage to these areas? Baddeley et al. (1988) have described a patient, P.V., who has a pure STM deficit. P.V. was able to learn meaningful paired associates in a familiar language. However, she was unable to learn to associate an unfamiliar word (in an unfamiliar language) with a familiar word in a familiar language, which is akin to learning a new vocabulary item. The fact that P.V. was a pure STM patient suggests that the critical damage in her case was to left inferior parietal cortex. This in turn indicates that this area of cortex does play a role in vocabulary acquisition. So, one part of the neural substrate commonly underlying verbal short-term memory and vocabulary acquisition is left inferior parietal cortex.

To summarize, the areas implicated in deficits in verbal short-term memory and vocabulary acquisition coincide broadly with the posterior neural correlates of phonological representation and storage, which comprise left temporo-parietal cortex and, in particular, the left inferior parietal lobule. This reliance of verbal short-term memory and word learning on the posterior portion of this substrate is evidenced by the span deficits of posterior lesion patients (Risse et al., 1984), and by the finding that in pure STM patients brain damage is primarily temporoparietal, and that such patients appear to have impaired vocabulary acquisition (Shallice & Vallar, 1990; Baddeley et al., 1988).

Rehearsal

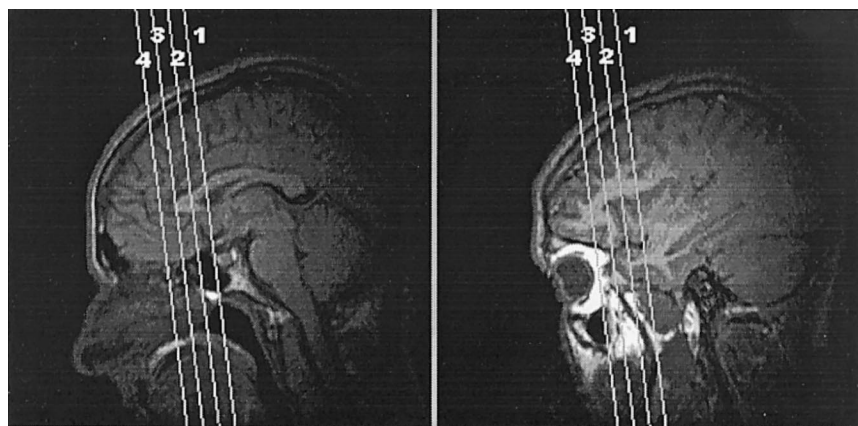
The speech planning/rehearsal component in our model maps onto Broca's area and other premotor cortices, the areas that subserve articulatory planning in speech.

One relevant line of evidence from neuropsychology indicates that digit span is compromised by certain articulatory disorders, but not by others. Thus, in dysarthric patients, muscular control for speech is impaired following subcortical and white matter lesions; such patients, however, have digit

span within the normal range and also exhibit word-length effects that suggest they are performing rehearsal (Vallar & Cappa, 1987; Bishop & Robson, 1989). In contrast, Waters et al. (1992) have examined patients with apraxia of speech, a condition involving impairment of speech-motor planning; these patients do have impaired digit span and also do not exhibit the effects that would be expected if they were performing rehearsal (in this respect, they are similar to normal subjects performing immediate serial recall under conditions of concurrent articulation). All the patients in this study had lesions that included Broca's area. To summarize, if Broca's area is spared, as is typical in dysarthria, rehearsal appears to be normal. If Broca's area is compromised, as is typical in apraxia of speech, rehearsal is impaired. These findings strongly suggest that Broca's area is in fact involved in articulatory rehearsal. This conclusion is also indirectly supported by the finding that Broca's area damage appears to impair patients' ability to increase articulation rate beyond a certain point (Blumstein, 1995); it seems likely that patients with such a deficit would also be impaired in rehearsal and in digit span.

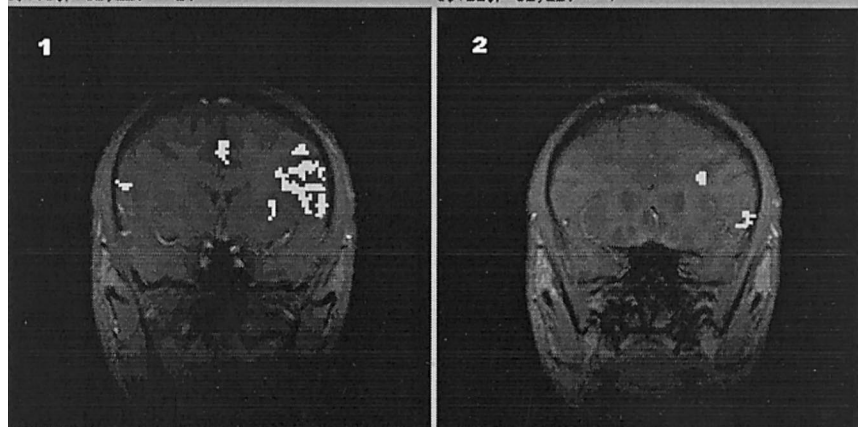
Further evidence comes from recent neuroimaging studies in which Broca's area has been found to be activated in verbal short-term memory tasks (Awh et al., 1994; Paulesu et al., 1993a, b; Petrides et al., 1993). In preliminary work of our own using functional magnetic resonance imaging (fMRI) we have obtained Broca's area activations during performance of a verbal working memory task that is thought to involve rehearsal (P. Gupta, J. D. Cohen, J. Jonides, and E. E. Smith, unpublished data). The task was a variant of the Continuous Performance Task (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956) designed to produce a continuous load on working memory. Figure 8 displays results from one subject. The figure highlights pixels that were significantly more activated in the experimental than in the control condition, using the *F* statistic, in one subject. As can be seen, a region corresponding to Broca's area is activated in the more posterior coronal images. We have observed similar activations in four of the six subjects tested in this study. The available evidence thus supports the hypothesis that rehearsal is dependent on the articulatory planning mechanisms subserved by Broca's area.

To summarize, the mapping of rehearsal onto Broca's area and other premotor speech cortices is supported by the finding that left anterior patients with apraxia of speech had impaired digit span and an apparent inability to rehearse (Waters et al., 1992). Anterior damage that does not cause articulatory planning deficits apparently does not impair verbal STM, as indicated by the preservation of STM abilities in dysarthrics (e.g., Vallar & Cappa, 1987) and in the patients with anterior lesions discussed by Risse, Rubens, and Jordan (1984), who were fluent and thus could not have had extensive damage to the premotor speech cortices. Both these groups of subjects had normal digit span; however, the vocabulary acquisition abilities of such pa-

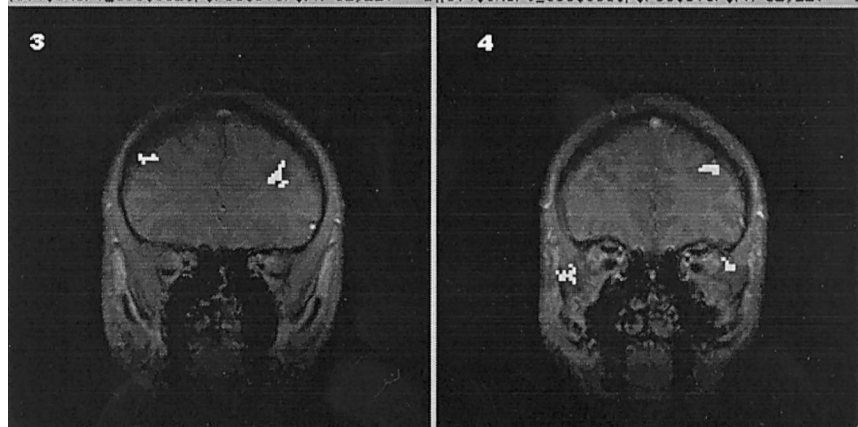


i.006,p 61,210 = 24

i.011,p 61,210 = 4



is04,short_isi.sl2.p.roi.over.Fm 61,210 = 2;is04,short_isi.sl3.p.roi.over.Fm 61,210 = 1



is04,short_isi.sl4.p.roi.over.Fm 61,210 = 1;is04,short_isi.sl5.p.roi.over.Fm 61,210 = 2

tients are unknown. These findings are exactly what would be expected if the rehearsal mechanisms of our cognitive model coincided with the neural subsystem we have identified as underlying articulatory planning. The mapping is also supported by neuroimaging studies.

Temporary Maintenance of Information

In the model, rehearsal processes entail the use of a queue of some kind. This component appears to map onto cortical areas that may be involved in the temporary maintenance of information, such as dorsolateral prefrontal cortex.

Referring again to the neuroimaging results in Fig. 8, it may be noted that the more anterior images show bilateral activation of regions corresponding to dorsolateral prefrontal cortex; this is consistent with a growing body of evidence from both neurophysiology and neuroimaging implicating dorsolateral prefrontal cortex (DLPFC) in working memory in human and nonhuman primates (e.g., Goldman-Rakic, 1987; Grasby et al., 1993; Jonides et al., 1993; Petrides et al., 1993; Blamire et al., 1993). Our study thus suggests that rehearsal and maintenance are both aspects of working memory and that DLPFC plays a role in maintenance of internal representations of stimuli that need to be held over temporal delays. In terms of the model, DLPFC maintains sequences of chunks.

It is further likely that there is a temporary storage loop between left temporo-parietal cortex and DLPFC. Functionally, such a loop represents the means of moving elements to and from the queue. The existence of connections between these areas is neurophysiologically plausible, as the posterior-most part of parietal association cortex does project to prefrontal cortex (Pandya & Yeterian, 1985).

In recent years, a number of important functional neural circuits have been identified and proposed to underlie various aspects of cognitive function in both human and nonhuman species. For example, Swerdlow and Koob (1987) have pointed to the importance of cortico-striato-pallido-thalamo-cortical loops in schizophrenia, mania, and depression, while a parallel model of such loops has been developed by Penney and Young (1983) to account for a variety of motor dysfunctions. In discussing the reciprocal connections between dorsolateral prefrontal cortex and posterior parietal cortex in pri-

FIG. 8. Preliminary results with fMRI in a verbal working memory task. Sagittal structural prescription images for one subject are shown in the upper row, one medial and one lateral. Statistical Parametric Maps (SPMs) representing performance of the cognitive task by the same subject are shown in the lower two rows, arrayed from posterior to anterior, corresponding to the locations indicated on the sagittal images. Broca's area is activated in posterior slices, and dorsolateral prefrontal cortex is activated bilaterally in anterior slices. Similar activations have been observed in other subjects.

mates, Goldman-Rakic (1987) has proposed that these areas and their connections compose a reverberating circuit for maintenance of the visuospatial representations needed in delayed-response performance.

It is therefore interesting that the system we have proposed incorporates two such loops. In one loop, frontal premotor areas constitute the anterior pole, responsible for articulatory planning, a necessary part of which is sequencing. Perisylvian areas of temporoparietal cortex compose the posterior pole, which is responsible for the formation of phonological representations. Mesulam (1990) has previously developed a similar conception of a linguistic premotor–temporoparietal system, although without focussing on the unique and key role of the frontal articulatory mechanisms in nonlinguistic cognitive phenomena such as verbal short-term memory. We also suggested that temporoparietal cortex and dorsolateral prefrontal cortex may be involved in a storage loop responsible for temporary maintenance of phonological information. In immediate serial recall, the parallel between this system and the prefrontal–parietal reverberating circuit proposed by Goldman-Rakic (1987) for maintenance of visuospatial representations is quite striking: the prefrontal–temporoparietal loop we have proposed potentially constitutes another such reverberating circuit.

Two points are worth noting about the functional circuits we have described. First, they are built around machinery specialized for core speech processing abilities (articulation and phonological representation). However, as we have proposed, the system participates in a variety of cognitive domains, providing an interesting example of the opportunistic utilization of neural structures in diverse ways. Second, given its specialization for speech, this system is likely to be uniquely human. This, in conjunction with its seemingly important role across cognitive domains, indicates that these functional neural circuits may be a central substrate of specifically human cognitive processing. Particularly interesting is the possibility that this system may underlie sequencing abilities: the role of the frontal premotor components may not be limited to sequencing articulatory movements into words, but may also subserve the sequencing of words into sentences in a way that influences syntax, as Mesulam (1990) has noted. More generally, this system may play a role in the formulation of ordered sequences of various kinds.

CONCLUSIONS

The earlier sections of this paper outlined a theoretical and computational approach to thinking about issues in vocabulary acquisition and in verbal short-term memory. We identified a specific set of phenomena that appear to relate verbal short-term memory and vocabulary acquisition and offered an integrated account of these phenomena. To recapitulate, our proposed model accounts for: (1) word and nonword repetition abilities; (2) the learning of new words via fast mapping; (3) the role of rehearsal in vocabulary

acquisition; (4) performance of ISR under various conditions and strategies; (5) capacity limitations and serial position effects in ISR; (6) the various phenomena suggesting the operation of an "articulatory" loop in ISR; (7) the role of rehearsal and chunking in ISR; and (8) differences between ISR of words and nonwords.

Most importantly, the model provides an account of how verbal short-term memory and vocabulary acquisition might be related. As conceptualized here, the relationship is at two levels. First, these abilities are related because they are dependent on the weights from the chunk layer to the phoneme layer and syllable template, i.e., on core phonological processing mechanisms. Second, they are related in their use of rehearsal and chunking. This constitutes the account of the correlation between performance in these two cognitive domains. It is worth emphasizing that this is a somewhat different conceptualization than one ascribing a causal role to verbal short-term memory (e.g., Gathercole & Baddeley, 1993). The present model views verbal short-term memory and vocabulary acquisition as involving common underlying mechanisms, without any implication that verbal short-term memory plays a causal role in vocabulary acquisition.

In summary, the model provides a theoretical framework within which a variety of questions can be posed regarding vocabulary acquisition, verbal short-term memory, and their relationship. The model as proposed here provides only a conceptual framework, rather than actual computer simulation results. However, it draws on existing computational work and is therefore a viable candidate for further development as a computational account. For the reasons discussed in the first section of this paper, the attempt to develop such a framework is important, and the present proposal represents a step toward developing one. Even in its present form, the model generates testable predictions and a framework for further inquiry.

The proposed model is composed of a phonological representational system and speech planning apparatus, which jointly underlie phenomena in the cognitive domains of vocabulary acquisition and verbal short-term memory. Following our computational specification of this system, we went on to consider how this cognitive system might relate to language processing mechanisms, adopting a cognitive neuroscientific strategy to examine this question.

We reviewed neuroanatomical, neurophysiological, neuropsychological, and neuroimaging evidence relating to the substrates of speech processing, which led us to a conceptualization that reconciles apparently contradictory data. In this view, lower "phonetic" levels of speech processing exhibit greater anterior-posterior specialization than do higher "phonological" levels of speech processing, which may constitute a more interactive anterior-posterior representational system. We further showed how the cognitive system we have proposed as underlying vocabulary acquisition and verbal short-term memory maps onto these neural substrates.

Viewed one way, our model offers an account of a wide range of behavioral phenomena relating to verbal short-term memory and vocabulary acquisition. Viewed another way, the model offers an account of what computations the neural substrates of verbal short-term memory and vocabulary acquisition might be performing. The present work thus offers an account linking brain and behavior. We are not aware of previous specifications of either the neural substrates of these cognitive phenomena or of their relationship to language processing mechanisms.

We view this work as the first presentation of an interesting new hypothesis linking brain, language, and verbal short-term memory. The computational ideas we have presented will of course have to be tested rigorously through computer simulation. The relationship of the central parts of our model to more peripheral aspects such as speech recognition will need to be elaborated. The mapping of the model onto the brain may well have to be revised as fresh evidence comes in, although we believe it is consistent with the data currently available.

These caveats notwithstanding, we believe the present work advances the discussion of brain and language, in that (a) it reconciles a large body of data at a behavioral, computational, and neural level, (b) it provides a theoretical framework for discussion, and (c) its claims are clearly laid out. Determining where the framework breaks down will therefore be as instructive as determining where its account is adequate. We therefore believe the present work is a step toward improved understanding of the relationship between brain, language, and verbal short-term memory.

REFERENCES

- Aaronson, D. 1968. Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology*, **76**, 129–140.
- Allport, A., Mackay, D. G., Prinz, W., & Scheerer, E. Eds. 1987. *Language perception and production: Relationships between listening, speaking, reading and writing*. New York: Academic Press.
- Amit, D. J., Sagi, D., & Usher, M. 1990. Architecture of attractor neural networks performing cognitive fast scanning. *Network*, **1**, 189–216.
- Au, T. K., & Glusman, M. 1990. The principle of mutual exclusivity in word learning: To honor or not to honor? *Child Development*, **61**, 1474–1490.
- Awh, E., Schumacher, E., Smith, E., Jonides, J., Koeppel, R., Minoshima, S., & Rowland, G. 1994. Investigation of verbal working memory using PET. In *Proceedings of the Inaugural Meeting of the Cognitive Neuroscience Society*. P. 28. [Abstract]
- Baddeley, A. D. 1986. *Working memory*. New York: Oxford Univ. Press.
- Baddeley, A. D. 1990a. The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge Univ. Press.
- Baddeley, A. D. 1990b. *Human memory*. Boston: Allyn and Bacon.
- Baddeley, A. D. 1992. Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, **4**, 281–288.
- Baddeley, A. D., & Hitch, G. 1974. Working memory. *The Psychology of Learning and Motivation*, **8**, 47–89.

- Baddeley, A. D., & Hitch, G. 1994. Developments in the concept of working memory. *Neuropsychology*, **8**, 485–493.
- Baddeley, A. D., Lewis, V., & Vallar, G. 1984. Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology A*, **36**, 233–252.
- Baddeley, A. D., Papagno, C., & Norris, D. 1991. Phonological memory and serial order: A sandwich for TODAM. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock*. Hillsdale, NJ: Erlbaum.
- Baddeley, A. D., Papagno, C., & Vallar, G. 1988. When long-term learning depends on short-term storage. *Journal of Memory and Language*, **27**, 586–595.
- Baddeley, A. D., Thomson, N., & Buchanan, M. 1975. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **14**, 575–589.
- Bairaktaris, D., & Stenning, K. 1992. A speech based connectionist model of human short term memory. In *Proceedings of the Fourteenth Conference of the Cognitive Society*. Hillsdale, NJ: Erlbaum.
- Bairaktaris, D., & Stenning, K. 1997. *Integrating phoneme and word levels in a model of immediate serial recall*. [manuscript in preparation].
- Baldwin, D. A., & Markman, E. M. 1989. Establishing word-object relations: A first step. *Child Development*, **60**, 381–398.
- Barlow, S. M., & Farley, G. R. 1989. Neurophysiology of speech. In D. P. Kuehn, M. L. Lemme, & J. M. Baumgartner (Eds.), *Neural bases of speech, hearing and language*. Boston, MA: Little, Brown.
- Bates, E., Thal, D., & Janowsky, J. 1988. Early language development and its neural correlates. In I. Rapin & S. Segalowitz (Eds.), *Handbook of neuropsychology*, Vol. 6, *Child neurology*. Amsterdam: Elsevier.
- Besner, D. 1987. Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology A*, **39**, 467–478.
- Bishop, D. V. M., & Robson, J. 1989. Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of “Articulatory Coding.” *Quarterly Journal of Experimental Psychology A*, **41**, 123–140.
- Bjork, E. L., & Healy, A. F. 1974. Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, **13**, 80–97.
- Blamire, A. M., McCarthy, G., Nobre, A. C., Puce, A., Hyder, F., Bloch, G., Phelps, E., Rothman, D., Goldman-Rakic, P., & Shulman, R. G. 1993. Functional magnetic resonance imaging of human pre-frontal cortex during a spatial memory task. *Proceedings of the Society of Magnetic Resonance in Medicine, Twelfth Annual Scientific Meeting*, **3**, 1413.
- Blumstein, S. E. 1990. Phonological deficits in aphasia: Theoretical perspectives. In A. Caramazza (Ed.), *Cognitive neuropsychology and neurolinguistics*. Hillsdale, NJ: Erlbaum.
- Blumstein, S. E. 1991. Phonological aspects of aphasia. In M. T. Sarno (Ed.), *Acquired aphasia*. New York: Academic Press. [2nd ed.]
- Blumstein, S. E. 1995. The neurobiology of the sound structure of language. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: MIT Press. Pp. 915–929.
- Borden, G. J. 1980. Use of feedback in established and developing speech. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*. New York: Academic Press. Vol. 3.
- Brown, G. D. A. 1989. A connectionist model of phonological short-term memory. In *Proceedings of the Eleventh Annual Conference of the Cognitive Society*. Hillsdale, NJ: Erlbaum. Pp. 572–579.
- Brown, G. D. A., & Hulme, C. 1995. Connectionist models of human short-term memory. In O. Omidvar (Ed.), *Progress in neural networks*. Norwood, NJ: Ablex. Vol. 3.
- Burgess, N. 1995. A solvable connectionist model of immediate recall of ordered lists. In G. Tesauro, D. Touretzky, & J. Alspector (Eds.), *Advances in neural information processing 7*. San Mateo, CA: Kaufmann.
- Burgess, N., & Hitch, G. J. 1992. Toward a network model of the articulatory loop. *Journal of Memory and Language*, **31**, 429–460.

- Callanan, M., & Markman, E. 1982. Principles of organization in young children's natural language hierarchies. *Child Development*, **53**, 1093–1101.
- Carey, S. 1978. The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.
- Carey, S., & Bartlett, E. 1978. Acquiring a single new word. In *Papers and Reports on Child Language Development*. Stanford University. Vol. 15, pp. 17–29.
- Chertkow, H., Bub, D., Evans, A., Meyer, E., & Marrett, S. 1997. *Cerebral activation during silent naming studied with positron emission tomography: A cortical correlate for subvocalization*. [Manuscript in preparation]
- Cohen, J. D., Forman, S. D., Casey, B. J., Servan-Schreiber, D., Noll, D. C., & Lewis, D. A. 1993. *Activation of prefrontal cortex in humans during a working memory task using functional MRI*. Society for Neuroscience Abstracts.
- Crystal, D. 1991. *A dictionary of linguistics and phonetics*. Cambridge, MA: Blackwell. [3rd ed.]
- Cuenod, C. A., Bookheimer, S., Pannier, L., Posse, S., Bonnerod, V., Turner, R., Gerrard, R., Frank, J. A., Zeffiro, T., & Lebihan, D. 1993. Functional imaging during word generation using a conventional MRI scanner. *Proceedings of the Society of Magnetic Resonance in Medicine, Twelfth Annual Scientific Meeting*, **3**, 1414.
- Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., & Frackowiak, R. 1992. The anatomy of phonological and semantic processing in normal subjects. *Brain*, **115**, 1753–1768.
- Dickinson, D. K. 1984. First impressions: Children's knowledge of words gained from a single exposure. *Applied Psycholinguistics*, **5**, 359–373.
- Dollaghan, C. 1985. Child meets word: "fast mapping" in preschool children. *Journal of Speech and Hearing Research*, **28**, 449–454.
- Dollaghan, C. 1987. Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders*, **52**, 218–222.
- Elman, J. L. 1990. Finding structure in time. *Cognitive Science*, **14**, 179–211.
- Evans, E. F. 1982. Functional anatomy of the auditory system. In H. B. Barlow & J. D. Mollon (Eds.), *The senses*. Cambridge, England: Cambridge Univ. Press. Chap. 14.
- Fox, P. T., Burton, H., & Raichle, M. E. 1987. Mapping human somatic sensory cortex with positron emission tomography. *Journal of Neurosurgery*, **67**, 34–43.
- Gathercole, S. E., & Baddeley, A. D. 1989. Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, **28**, 200–213.
- Gathercole, S. E., & Baddeley, A. D. 1990. Phonological memory deficits in language-disordered children: Is there a causal connection? *Journal of Memory and Language*, **29**, 336–360.
- Gathercole, S. E., & Baddeley, A. D. 1993. *Working memory and language*. Hillsdale, NJ: Erlbaum.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. 1991. The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, **12**, 349–367.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. 1994. The children's test of nonword repetition: A test of phonological working memory. *Memory*, **2**, 103–127.
- Geschwind, N. 1979. Specializations of the human brain. *Scientific American*, **241**(3), 190–199.
- Glasspool, D. W. 1997. *Serial recall of words and nonwords in an articulatory loop model*. [Manuscript in preparation]
- Gnadt, J. W., & Anderson, R. A. 1988. Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research*, **70**, 216–220.
- Goldman, R., Fristoe, M., & Woodcock, R. 1974. *Technical manual for Goldman-Fristoe-Woodcock Auditory Skills Test Battery*. Circle Pines, MN: American Guidance Service.

- Goldman-Rakic, P. S. 1987. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology*. Bethesda, MD: American Physiological Society. Vol. 5.
- Gracco, V. L., & Abbs, J. H. 1987. Programming and execution processes of speech movement control: Potential neural correlates. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.
- Grasby, P. M., Frith, C. D., Friston, K. J., Bench, C., Frackowiak, R. S. J., & Dolan, R. J. 1993. Functional mapping of brain areas implicated in auditory-verbal memory function. *Brain*, **116**, 1–20.
- Grossberg, S. 1969. Some networks that can learn, remember, and reproduce any number of complicated space-time patterns. *Journal of Mathematics and Mechanics*, **19**, 53–91.
- Grossberg, S. 1978. A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology*. New York: Academic Press. Vol. 5.
- Grossberg, S. 1986. The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines*, Vol. 1, *Speech perception*. New York: Academic Press.
- Grossberg, S. 1987. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, **11**, 23–63.
- 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.
- Gulick, W. L., Gescheider, G. A., & Frisina, R. D. 1989. *Hearing: Physiological acoustics, neural coding, and psychoacoustics*. New York: Oxford Univ. Press.
- Gupta, P. 1996. Immediate Serial Memory and Language Processing: Beyond the Articulatory Loop. Technical Report No. CS-96-02. Urbana: Beckman Institute, Cognitive Science Group.
- Gupta, P., & MacWhinney, B. 1993. Is the phonological loop articulatory or auditory? In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum. Pp. 510–515.
- Gupta, P., & MacWhinney, B. 1995. Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language*, **34**, 63–88.
- Hanley, R. J., & Broadbent, C. 1987. The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology*, **78**, 287–297.
- Harrison, R. V., & Hunter-Duvar, I. M. 1989. Neuroanatomy of hearing. In D. P. Kuehn, M. L. Lemme, & J. M. Baumgartner (Eds.), *Neural bases of speech, hearing and language*. Boston, MA: Little, Brown.
- Hartley, T., & Houghton, G. 1996. A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, **35**, 1–31.
- Hinke, R. M., Hu, X., Stillman, A. E., & Ugurbil, K. 1993. The use of multislice functional MRI during internal speech to demonstrate the lateralization of language function. *Proceedings of the Society of Magnetic Resonance in Medicine, Twelfth Annual Scientific Meeting*, **1**, 63.
- Holdgrafer, G., & Sorensen, P. 1984. Informativeness and lexical learning. *Psychological Reports*, **54**, 75–80.
- 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*. New York: Academic Press.
- Houghton, G. 1993. Inhibitory control of neurodynamics: Opponent mechanisms in sequencing and selective attention. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology*. New York: Academic Press.
- Houghton, G. 1994. Spelling and serial recall: Insights from a competitive queueing model.

- In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of normal and disturbed spelling*. New York: Wiley.
- Howard, D., & Franklin, S. 1990. Memory without rehearsal. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge Univ. Press.
- Howard, D., & Franklin, S. 1993. Dissociations between component mechanisms in short-term memory: Evidence from brain-damaged patients. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV*. Cambridge, MA: MIT Press.
- Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., & Frackowiak, R. 1992. The cortical localization of the lexicons. *Brain*, **115**, 1769–1782.
- Hulme, C., Maughan, S., & Brown, G. D. A. 1991. Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, **30**, 685–701.
- Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. 1993. Spatial working memory in humans as revealed by PET. *Nature*, **363**, 623–625.
- Jordan, M. I. 1986. *Serial order: A parallel distributed processing approach*. Report 8604, Institute for Cognitive Science, University of California, San Diego, La Jolla, CA.
- Kandel, E. R., & Schwartz, J. H. Eds. 1985. *Principles of neural science*. New York: Elsevier. [2nd ed.]
- Kean, M. L. Ed. 1985. *Agrammatism*. New York: Academic Press.
- Keefe, K. A., Feldman, H. M., & Holland, A. L. 1989. Lexical learning and language abilities in preschoolers with perinatal brain damage. *Journal of Speech and Hearing Disorders*, **54**, 395–402.
- Keenan, J. M., & MacWhinney, B. 1987. Understanding the relationship between comprehension and production. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of language production*. Norwood, NJ: Ablex.
- Keller, E. 1987. The cortical representation of motor processes of speech. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.
- Keller, E., & Gopnik, M. Eds. 1987. *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.
- Kertesz, A., Harlock, W., & Coates, R. 1979. Computer tomographic localization, lesion size, and prognosis in aphasia and nonverbal impairment. *Brain and Language*, **8**, 34.
- Kohonen, T. 1984. *Self-organization and associative memory*. Berlin: Springer-Verlag.
- Kuehn, D. P., Lemme, M. L., & Baumgartner, J. M. Eds. 1989. *Neural bases of speech, hearing and language*. Boston, MA: Little, Brown.
- Kuhl, P. K. 1991. Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S. E. Brauth, W. S. Hall, & R. J. Dooling (Eds.), *Plasticity of development*. Cambridge, MA: MIT Press.
- Kuhl, P. K., & Meltzoff, A. N. 1982. The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.
- Kupfermann, I. 1985. Hemispheric asymmetries and the cortical localization of higher cognitive and affective functions. In E. R. Kandel & J. H. Schwartz (Eds.), *Principles of neural science*. New York: Elsevier. 2nd ed.
- Lassen, N. A., Ingvar, D. H., & Skinhoj, E. 1978. Brain function and blood flow. *Scientific American*, 239–250.
- Lassen, N. A., & Larsen, B. 1980. Cortical activity in the left and right hemispheres during language-related brain functions. *Phonetica*, **37**, 27–37.
- Leonard, L. B., Schwartz, R. G., Morris, B., & Chapman, K. 1981. Factors influencing early lexical acquisition: Lexical orientation and phonological composition. *Child Development*, **52**, 882–887.
- Liepmann, H. 1915. Diseases of the brain. In C. W. Barr (Ed.), *Curschmann's textbook on nervous diseases*. Philadelphia: Blakiston. Vol. 1.
- Mackay, D. G., Allport, A., Prinz, W., & Scheerer, E. 1987. Relationships and modules within

- language perception and production: An introduction. In A. Allport et al. (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing*. New York: Academic Press.
- Markman, E. 1984. The acquisition and hierarchical organization of categories by children. In C. Sophian (Ed.), *Origins of cognitive skills*. Hillsdale, NJ: Erlbaum.
- Markman, E. 1989. *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- Markman, E. M., & Hutchinson, J. E. 1984. Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, **16**, 1–27.
- Mateer, C. A. 1989. Neural correlates of language function. In D. P. Kuehn et al. (Eds.), *Neural bases of speech, hearing and language*. Boston, MA: Little, Brown.
- Mayeux, R., & Kandel, E. R. 1985. Natural language, disorders of language, and other localizable disorders of cognitive functioning. In E. R. Kandel & J. H. Schwartz (Eds.), *Principles of neural science*. New York: Elsevier.
- McAdam, D. W., & Whitaker, H. A. 1971. Language production: Electroencephalographic localization in the normal human brain. *Science*, **172**, 499–502.
- McClelland, J. L., & Elman, J. L. 1986. Interactive processes in speech perception: The TRACE model. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing*, Volume 2, *Psychological and Biological Models*. Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. 1988. *Exploration in parallel distributed processing*. Cambridge, MA: MIT Press.
- McNaughton, B. L., & Morris, R. G. M. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences*, **10**(10), 408–415.
- Merriman, B., & Schuster, J. 1991. Young children's disambiguation of name reference. *Child Development*, **62**, 1288–1301.
- Merriman, W., & Bowman, L. 1989. The mutual exclusivity bias in word learning. *Monographs of the Society for Research in Child Development*, **54**(3–4), 1–129.
- Mervis, C. 1984. Early lexical development: The contributions of mother and child. In C. Sophian (Ed.), *Origins of cognitive skills*. Hillsdale, NJ: Erlbaum.
- Mervis, C., Golinkoff, R., & Bertrand, J. 1994. Two-year-olds readily learn multiple labels for the same basic-level category. *Child Development*, **65**, 1163–1177.
- Mervis, C., & Pani, J. 1980. Acquisition of basic object categories. *Cognitive Psychology*, **12**, 496–522.
- Mesulam, M. M. 1990. Large-scale neurocognitive networks and distributed processing for attention, language and memory. *Annals of Neurology*, **28**, 597–613.
- Miikkulainen, R. 1990. A distributed feature map model of the lexicon. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- 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.
- Mohr, J. P. 1976. Broca's area and Broca's aphasia. In H. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics*. New York: Academic Press. Vol. 1.
- Monsell, S. 1987. On the relation between lexical input and output pathways for speech. In A. Allport et al. (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing*. New York: Academic Press.
- Naeser, M. A., & Hayward, R. W. 1978. Lesion localization in aphasia with cranial-computed tomography and Boston diagnostic aphasia exam. *Neurology*, **28**, 545.
- Nelson, K. E., & Bonvillian, J. D. 1973. Concepts and words in the 18-month-old: Acquiring concept names under controlled conditions. *Cognition*, **2**, 435–450.
- Ojemann, G. A. 1983. Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences*, **6**, 189–230.

- Oviatt, S. 1980. The emerging ability to comprehend language: An experimental approach. *Child Development*, **51**, 97–106.
- Oviatt, S. 1982. Inferring what words mean: Early development in infants' comprehension of common object names. *Child Development*, **53**, 274–277.
- Pandya, D. N., & Yeterian, E. H. 1985. Architecture and connections of cortical association areas. In A. Peters & E. G. Jones (Eds.), *Association and auditory cortices*, Vol. 4, *Cerebral cortex*. New York: Plenum.
- Papagno, C., Valentine, T., & Baddeley, A. D. 1991. Phonological short-term memory and foreign-language learning. *Journal of Memory and Language*, **30**, 331–347.
- Paulesu, E., Bottini, G., Bench, C. J., Grasby, P. M., Frith, C. D., & Frackowiak, R. S. J. 1993a. The double anatomical dissociation of verbal and visual short-term memory: A PET activation study. *Society for Neuroscience Abstracts*, **19**, 413.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. 1993b. The neural correlates of the verbal component of working memory. *Nature*, **362**, 342–345.
- Penney, J. B., & Young, A. B. 1983. Speculations on the functional anatomy of basal ganglia disorders. *Annual Review of Neuroscience*, **6**, 73–94.
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. 1989. Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, **1**(2), 153–170.
- Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. 1993. Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Sciences USA*, **90**, 878–882.
- Richman, H. B., & Simon, H. A. 1997. *EPAM simulations of short-term memory*. [Manuscript in preparation]
- Risse, G. L., Rubens, A. B., & Jordan, L. S. 1984. Disturbance of long-term memory in aphasic patients: A comparison of anterior and posterior lesions. *Brain*, **107**, 605–617.
- Rosvold, K. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. 1956. A continuous performance test of brain damage. *Journal of Consulting Psychology*, **20**, 343–350.
- Rueckert, L., Appollonio, I., Grafman, J., Johnson Jr., R., Le Bihan, D., & Turner, R. 1993. Functional activation of left frontal cortex during covert word production. *Proceedings of the Society of Magnetic Resonance in Medicine, Twelfth Annual Scientific Meeting*, **1**, 60.
- Salame, P., & Baddeley, A. D. 1982. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, **21**, 150–164.
- Schwartz, R. G., & Leonard, L. 1984. Words, objects and actions in early lexical acquisition. *Journal of Speech and Hearing Research*, **27**, 119–127.
- Seldon, H. L. 1985. The anatomy of speech perception. In A. Peters & E. G. Jones (Eds.), *Association and Auditory Cortices*, Vol. 4, *Cerebral cortex*. New York: Plenum.
- Service, L. 1992. Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology A*, **45**, 21–50.
- Shallice, T., & Vallar, G. 1990. The impairment of auditory-verbal short-term storage. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge Univ. Press.
- Studdert-Kennedy, M. 1986. Development of the speech perceptuomotor system. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of Early Speech*. New York: Stockton Press.
- Swerdlow, N. R., & Koob, G. F. 1987. Dopamine, schizophrenia, mania, and depression: Toward a unified hypothesis of cortico-striato-pallido-thalamic function. *Behavioral and Brain Sciences*, **10**, 197–245.
- Treiman, R., & Danis, C. 1988. Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 145–152.

- Tyler, L. K. 1992. *Spoken language comprehension*. Cambridge, MA: MIT Press.
- Vallar, G., & Cappa, S. F. 1987. Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology*, **4**(1), 55–78.
- Vallar, G., & Shallice, T. Eds. 1990. *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge Univ. Press.
- Wang, D., & Arbib, A. M. 1991. A neural model of temporal sequence generation with interval maintenance. In *Proceedings of the Thirteenth Conference of the Cognitive Society*. Hillsdale, NJ: Erlbaum. Pp. 944–948.
- Waters, G. S., Rochon, E., & Caplan, D. 1992. The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language*, **31**, 54–73.
- Yost, W. A., & Gourevitch, G. Eds. 1987. *Directional hearing*. New York: Springer-Verlag.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science*, **256**, 846–849.
- Zhang, G., & Simon, H. A. 1985. STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition*, **13**, 193–201.