

A NEURAL NETWORK MODEL OF VERBAL WORKING MEMORY BASED ON TRANSITORY ACTIVATION PATTERNS*

Shane T. Mueller and David E. Meyer
University of Michigan

Sponsored by the U.S. Office of Naval Research

*Poster Presented at the Eighth Annual
Meeting of the Cognitive Neuroscience
Society, March 25-27, 2001.

Abstract

Traditionally, cognitive neuroscientists have represented long-term memory in terms of the structure of a neural network's connections and short-term memory in terms of the patterns of activation across the network (e.g., Hebb, 1949; Caianiello, 1961). However, recent neural-network models of short-term verbal working memory (VWM) have used modifiable structural connections to encode item and order information (Hartley & Houghton, 1996; Burgess & Hitch, 1999). In these models, words are stored by changing the connection weights between linguistic units, and phenomena related to VWM are thus modeled with long-term memory structures and mechanisms. Although it may be possible for neural connections to change rapidly (e.g., see Jensen et al., 1996), this latter approach to modeling VWM does not appear to be motivated by neurobiology. Consequently, we have formulated a neural-network model of VWM that stores and maintains item and order information as a pattern of activation within a fixed-structure network whose connections remain constant. Performance with this network is affected by many of the standard factors that influence VWM performance, such as phonological similarity, articulatory duration, and word frequency. Our work demonstrates that the embodiment of short-term memory based on patterns of network activation is feasible, parsimonious, and merits more investigation. Furthermore, from our work it appears that both the biology and psychology of specific mental processes must be understood more fully before neural networks can be deemed "neurally plausible".

Background and Present Goals

- Hebb (1949) tried to determine whether neural coding depends on:
 - (1) the activation of specific cellular units (e.g., Hull);
 - (2) a pattern of neural activation whose locus is unimportant (cf. Lashley, Kohler).
- Based on his conclusions, Hebb formulated neural network models of type (1) that coded information in terms of cellular activation.
- Following Hebb, recent neural network models have used representations based on the activation of specific units, and associations between these units are formed by increasing the strength of their interconnections.
- To account for short-term “working” memory phenomena, these recent models implement short-term memory storage via mechanisms and structures like those traditionally associated with long-term memory.
- In contrast, the present theoretical research has two new goals:
 - (1) to formulate a neural network model that codes information in a pattern of activation whose locus is unimportant;
 - (2) to account for short-term memory phenomena with a fixed-structure network that does not encode short-term memory information in LTM structures.

Proposed Verbal WM Network Model

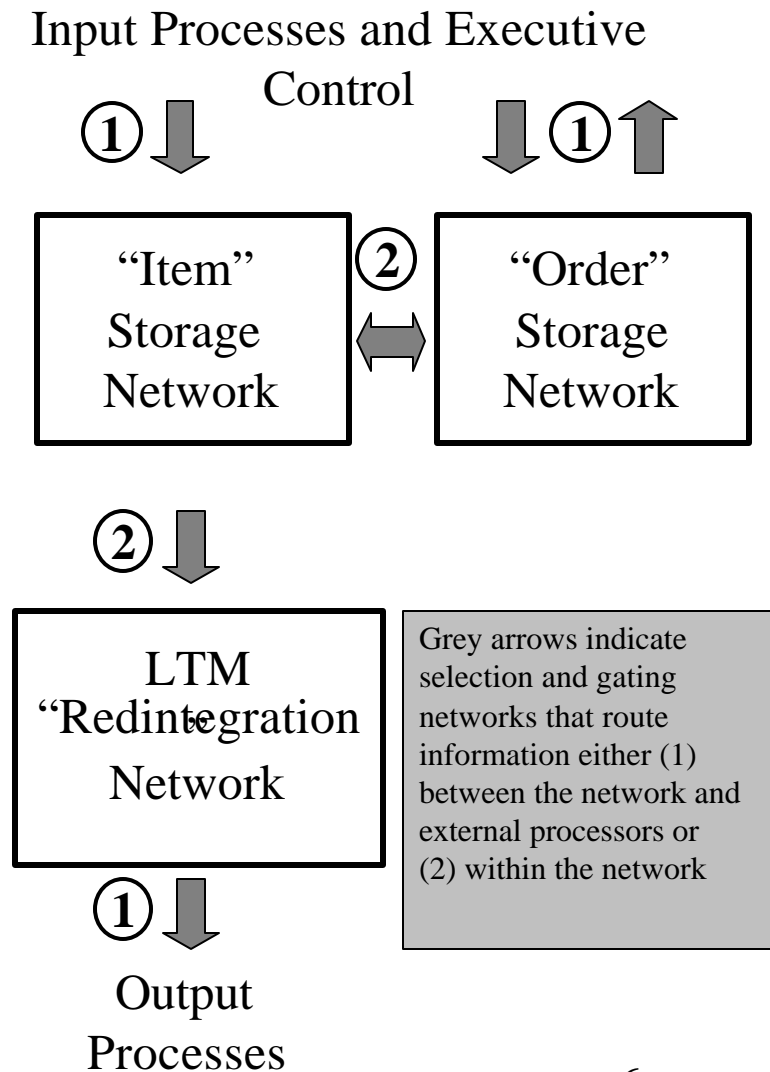
- We propose a network model of verbal working memory that encodes and stores short-term verbal information in a fixed-structure network whose connection weights are constant.
- Our proposed model differs fundamentally in several ways from other recent models.
- For our model, short-term information is stored exclusively via recurrent activation patterns in a fixed-structure network.
- On the other hand, in recent neural network models of “short-term” verbal working memory (VWM), performance is usually modeled through changes in the morphological structure of a network, i.e., “connection weights” (e.g., Burgess & Hitch, 1999; Lisman & Ideart, 1995; Hartley & Houghton, 1996; Page & Norris, 1998).

Virtues of Proposed Model

- Our proposed model implements a neural-network model of short-term VWM without using LTM “learning” mechanisms.
- Verbal information is stored temporarily as a recurrent pattern of activation, and it is routed within the network using gating mechanisms.
- The model acknowledges that there are distinct storage mechanisms and characteristics for verbal “Item” and “Order” information.
- The model provides new interpretations and predictions about patterns of “activation” found in neuroimaging studies (e.g., Smith et al., 1998).

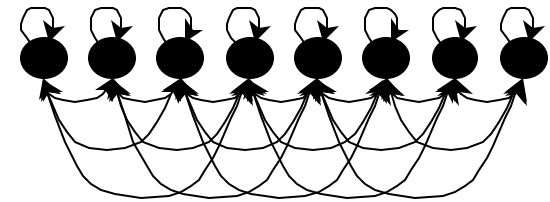
Global Structure of Network

- Our network model is composed of three major interactive modules:
 - the “**Item**” storage network;
 - the “**Order**” storage network;
 - the LTM “**Redintegration**” network.
- Modules are connected by **selection** and **gating** networks that route information between various regions.
- “Executive” control processes external to the network use it as a system for encoding, maintaining, and retrieving short-term verbal information for tasks such as serial list recall.



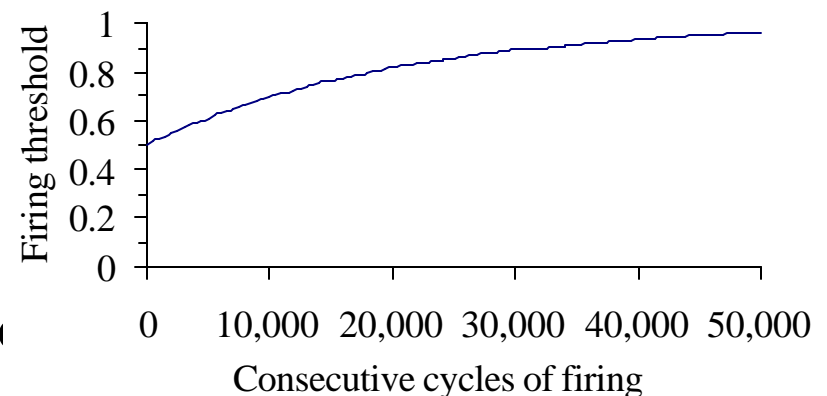
Structure of Network Buffers

- “Phonological” and “location” storage buffers exist within the “Item” and “Order” networks, respectively.
- These buffers consist of many interconnected neuron-like units.
- Units are connected to themselves and other nearby units.
- Processing proceeds in successive cycles.
- During each cycle, a unit is either “on” or “off”.
- If the activation of a unit exceeds a threshold, the unit fires on the next cycle.
- When a unit fires, its threshold increases, making the unit less likely to fire again.



Above: The low-level interconnections for a region in the network buffer. Units are connected to themselves and nearby units.

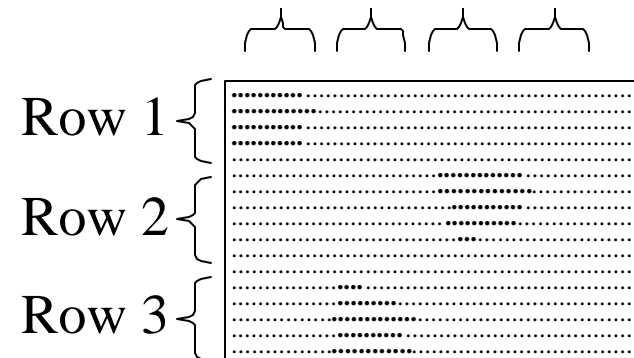
Below: The threshold for a single unit firing again increases on each cycle that the unit fires, and decreases on each cycle that the unit does not fire, ranging from a minimum of 0.5 to a maximum of 1.0. The figure shows the functional relationship between threshold and consecutive cycles of firing.



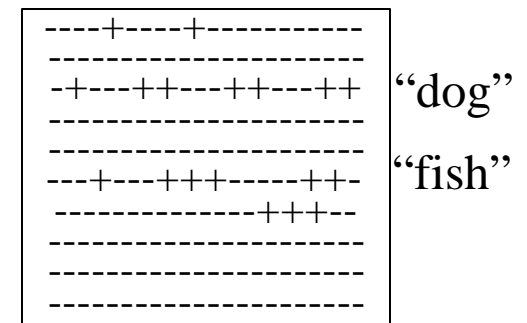
Representation of Words in Buffers

- The “Item” information for a word is coded in terms of its phonological features.
- The “Order” information for a word is coded as an index to the location in the storage buffer for the next word.
- Individual phonological features are coded by the activation of many contiguous units.
- Individual words are represented as a single row of features, coded by the activation of many contiguous rows of units.
- Patterns of activation, independent of their location in the buffer, embody information about an encoded word.

Units in the Storage Buffer
Individual Features

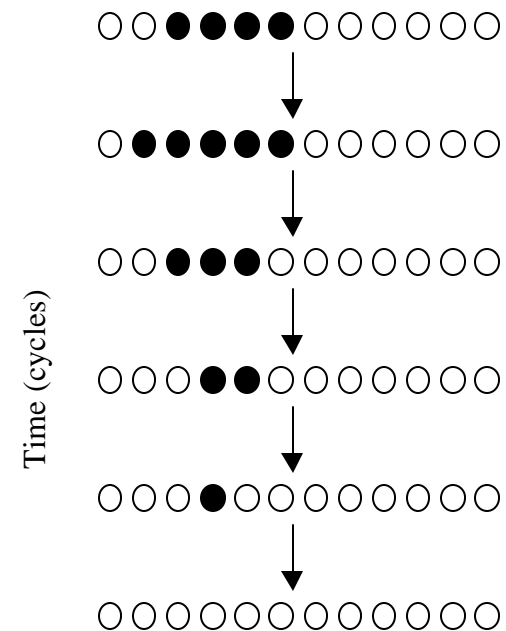


Typical Activation Patterns



Activation Dynamics in Storage Buffers

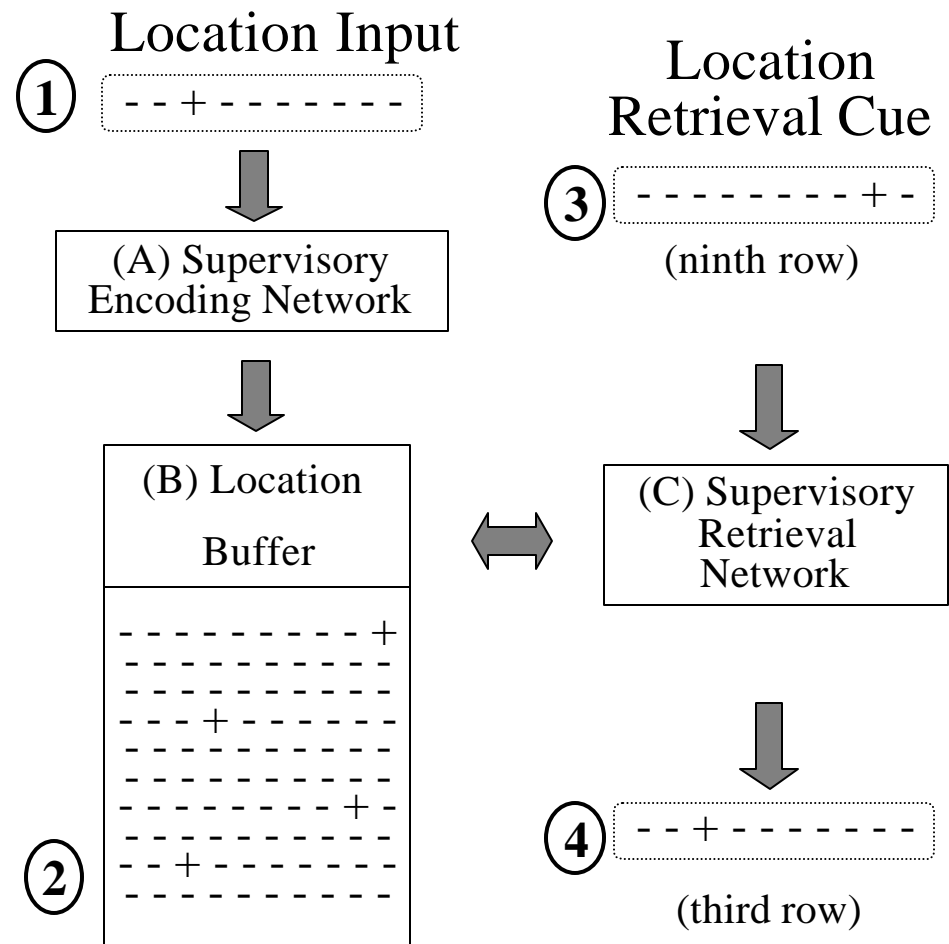
- Verbal information is coded as a persistent pattern of activation.
- Local excitatory connections encourage active blocks of units to remain active, and inactive blocks to remain inactive.
- Activation decays with time, because either:
 - (1) Incoming stochastic activation dips below the activation threshold;
 - (2) The slowly increasing threshold exceeds typical levels of activation.



Typical operation of storage network. When a small block of units in the buffer is activated, the activation is maintained in this region for a period of time. Eventually, because of a slowly increasing threshold, the activation will fail to be propagated to the next cycle.

“Order” Storage Network

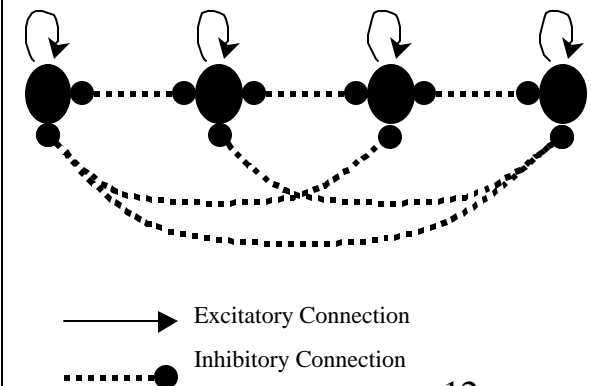
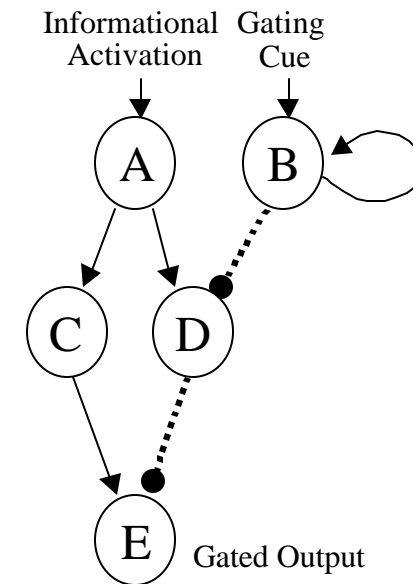
- Basic structure of “Order” network is similar to “Item” network.
- Location buffer (B) stores serial order information through the activation of a block of units that indicates the next item’s row in the location buffer.
- Supervisory encoding network (A) places serial order information for new items in inactive regions of the location buffer.
- Supervisory retrieval network (C) selects information from cued locations of the location buffer.



Above diagram depicts the operation of the “Order” storage network. The order information indicating that the next item will be stored in the third row is presented (1) and encoded in the ninth row (2). Later, the information in the ninth row is cued for retrieval (3), and retrieved from the network (4).

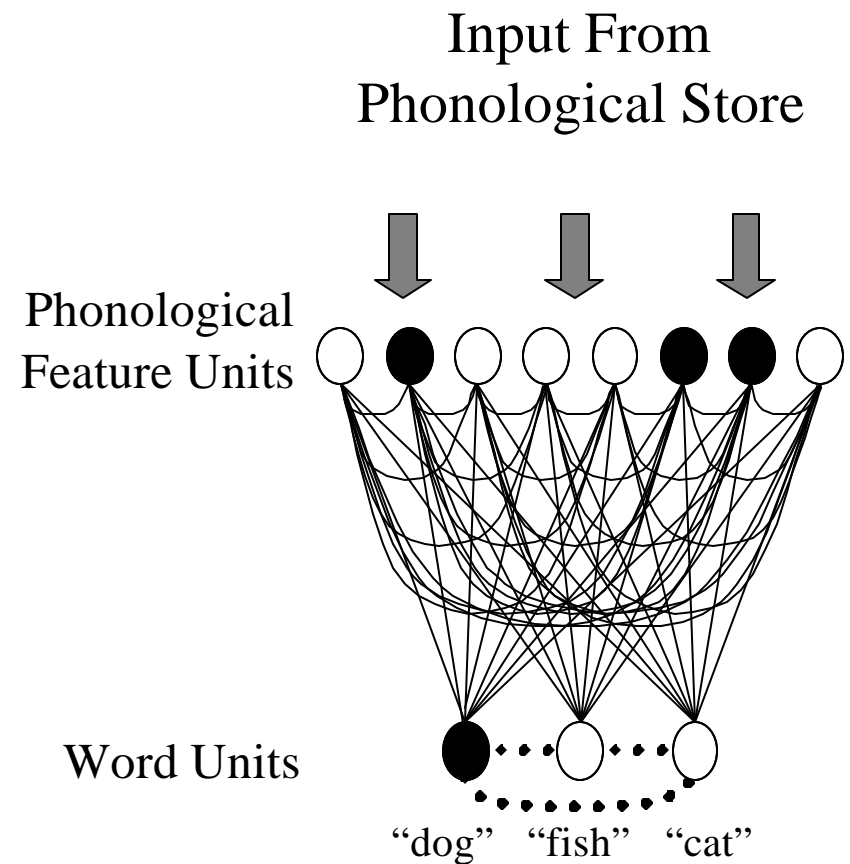
Structure of Gating Mechanisms

- Adjacent to the network buffers are gating mechanisms that select information and route it through the network.
- The top right panel shows a **Gate**, in which:
 - Normally, activity of unit A does not affect unit E because activity of unit D counteracts that of C.
 - When the gating cue at unit B is activated, it “opens” the gate by activating itself and inhibiting unit D.
 - When the gate is opened by unit B, activity in unit A passes through unit C and activates unit E.
 - Eventually, the gate “closes” because the threshold of unit B increases, stopping B from firing.
- Bottom panel shows a **winner-take-all selector**:
 - When a single unit fires, it inhibits other units.
 - Eventually, a single unit fires exclusively.



Structure of LTM Network

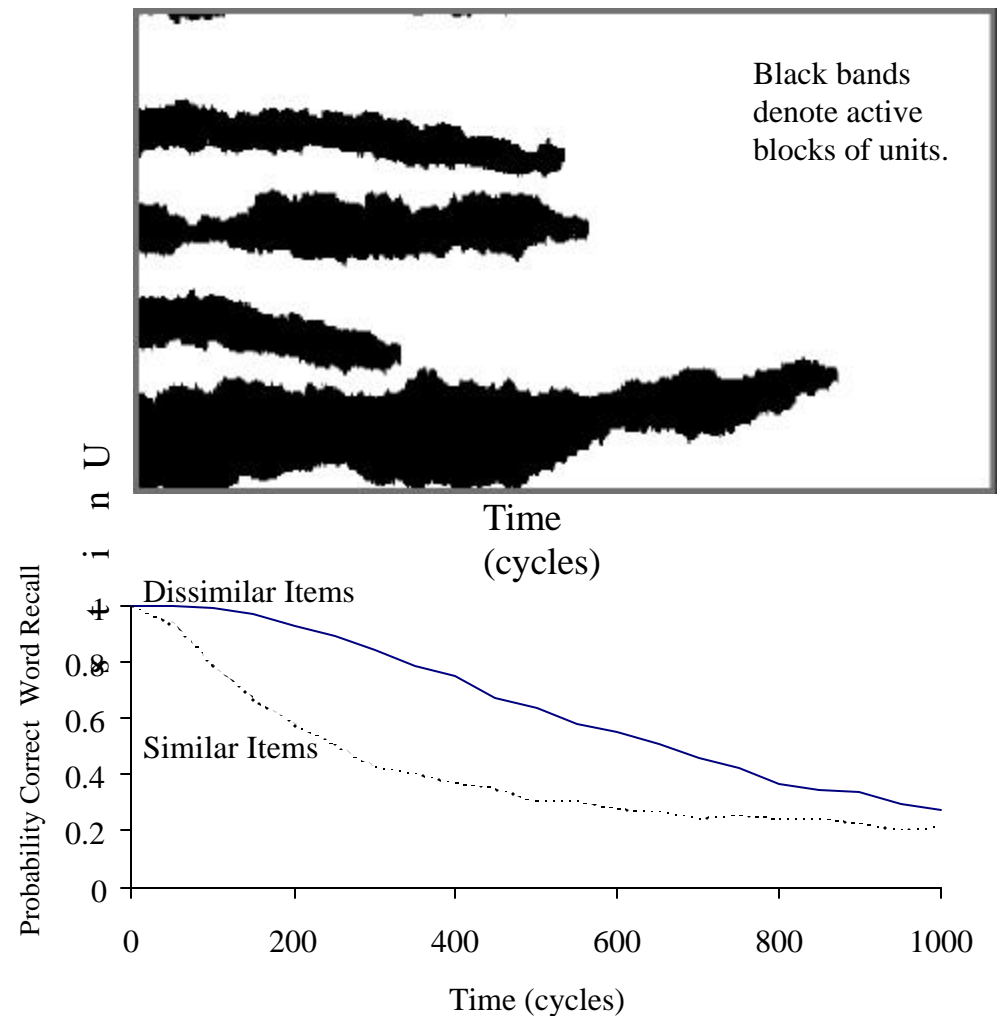
- Degraded information from the phonological buffer is “redintegrated” by a long-term memory attractor network.
- The attractor network has units whose interconnections store all relevant words (e.g., all words used during a memory-span task).
- For each stored word, there is a stable pattern of activation in the network.
- Redintegration is poorer for patterns that are not well-learned, such as those associated with infrequent words or non-words.



Properties of “Item” Information

- “Item” information is stored in the phonological buffer.
- As time passes, activation patterns for phonological features decay, but an entire word can be reconstructed based on residual activation.
- Together, the “Item” and LTM networks produce observable decay functions (probability of correct recall versus time) for similar and dissimilar word sets.
- Decay distributions are approximately log-normal.

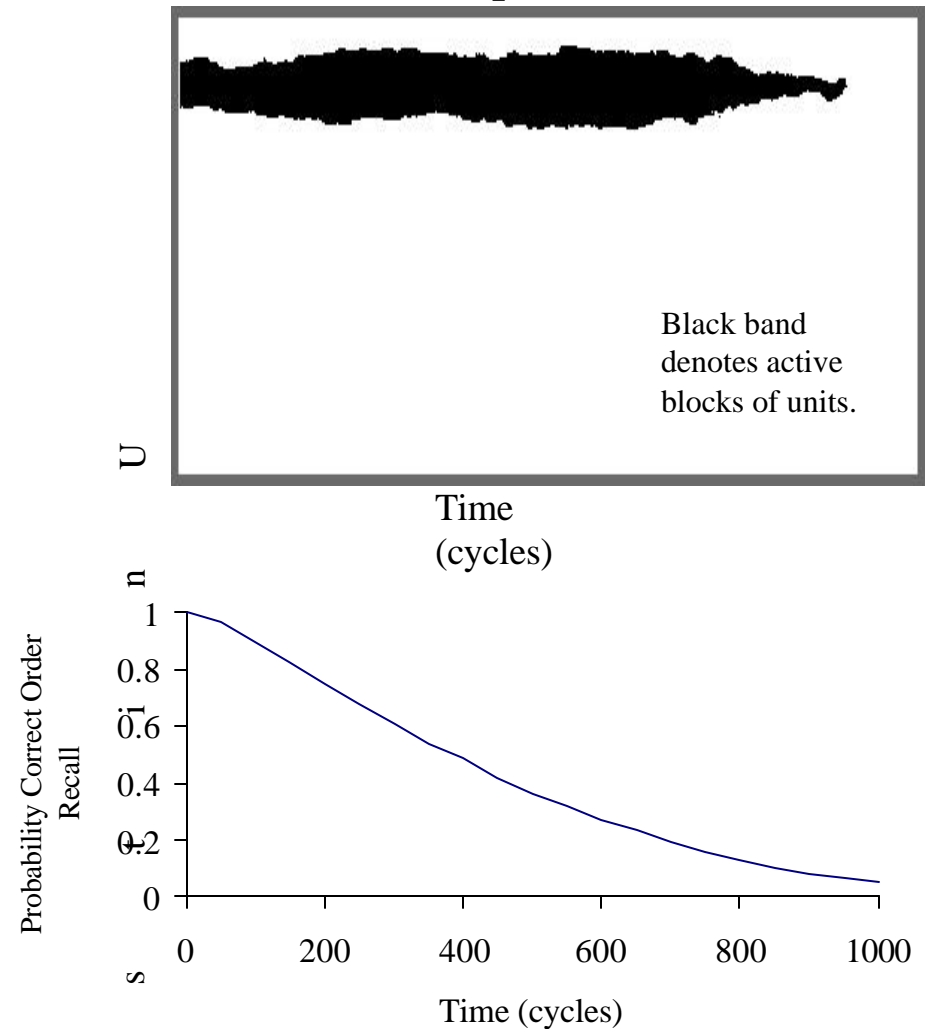
Example Trial



Properties of “Order” Information

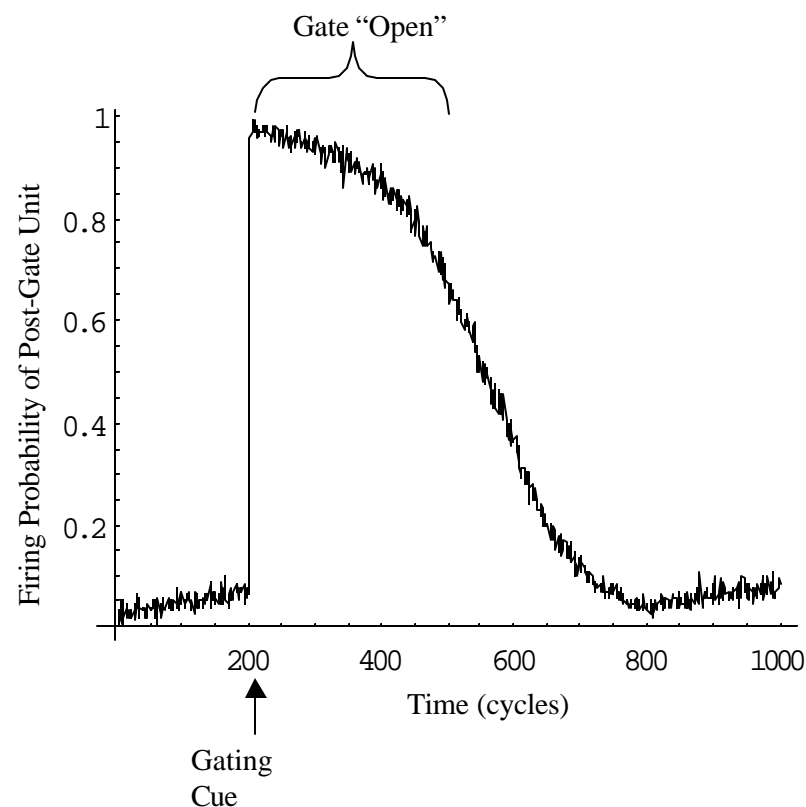
- Serial order information is stored in the location buffer as a single band of activation.
- This activation band eventually fails to propagate itself, “decaying” with time.
- Although “Order” information decays like the individual features of the “Item” information, “Order” information is less robust because it cannot be reintegrated.

Example Trial



Performance of Gating Network

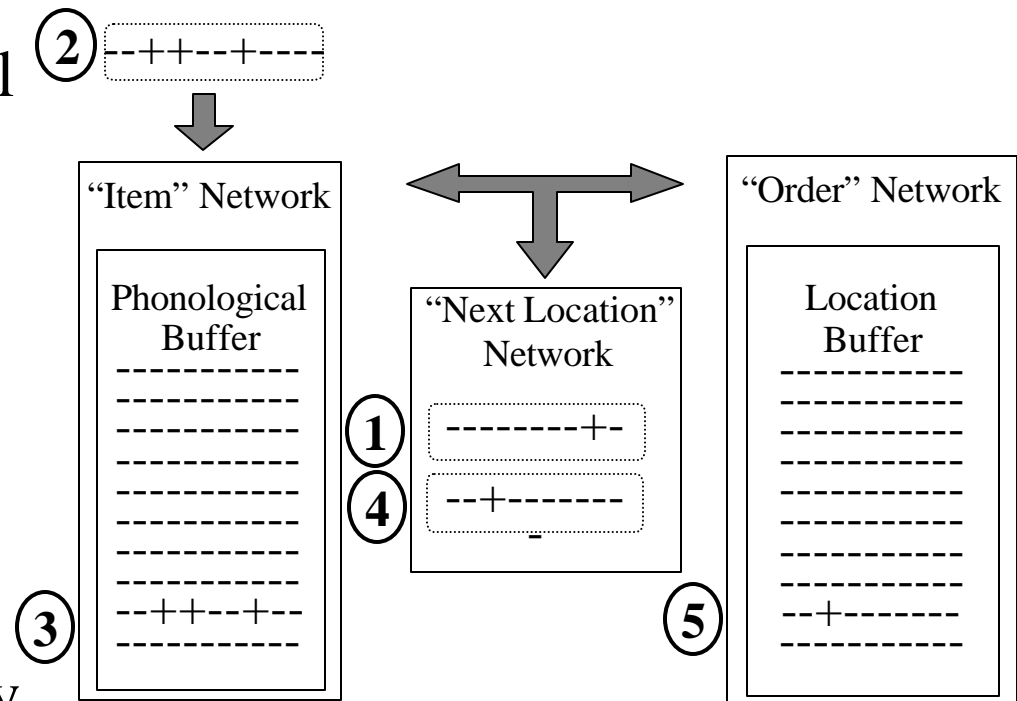
- When a gate is opened, activation passes through immediately.
- The gate remains open until the gating unit (B) fails to reactivate itself.
- The length of time that the gate remains open varies stochastically.
- This time depends on the rate at which the gating unit's activation threshold increases.



The operation of gating mechanism is shown above. Here, the informational (pre-gate) unit is active throughout the simulation. When a gating cue is presented for a single cycle, the gate “opens”, allowing activation to pass through the gate for a period of time. Eventually, the gate “closes”, because of a slowly increasing activation threshold.

Network Operation: Storage

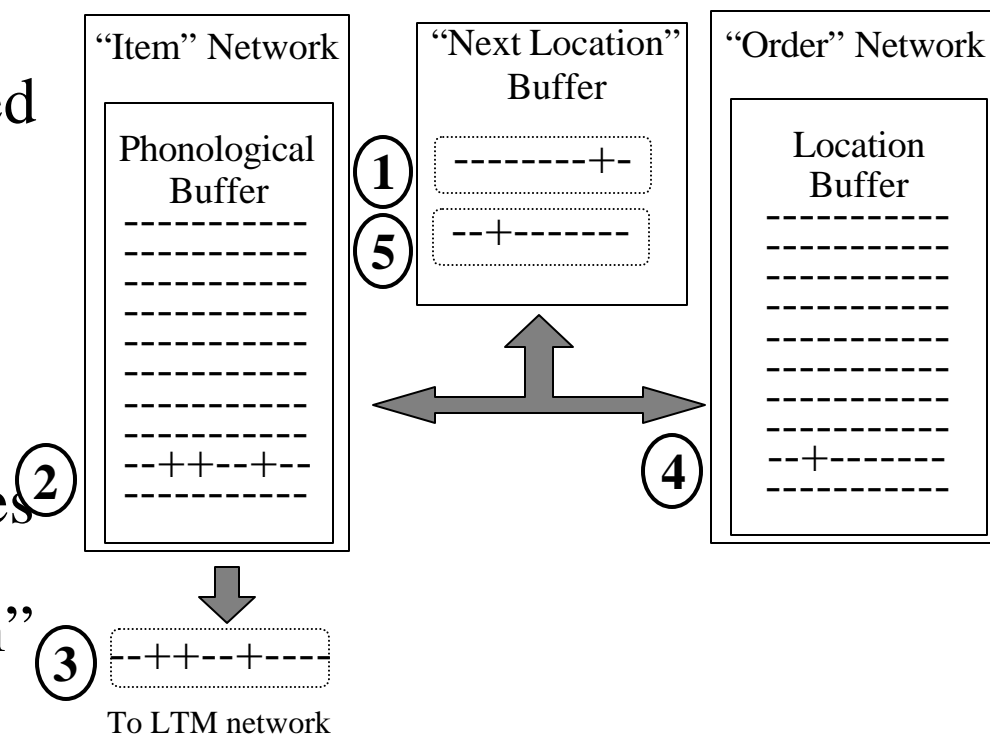
- The “Item” network stores information in a phonological buffer, and the “Order” network stores information in a location buffer.
- Corresponding rows of the phonological and location buffers contain information about the same word.
- Storage operations in the “Item” and “Order” networks are coordinated by supervisory gating networks, such as the “Next Location” network.
- The “Next Location” network selects an empty row of the phonological buffer, using a “winner-take-all” mechanism.



The above diagram shows the operation of the “Item” and “Order” networks during storage. This operation is modulated through a “Next Location” supervisory network. At point (1), the “Next Location” network indicates that the next encoded item will be stored in the ninth row. Next, a new item (2) is encoded into the “Item” network (3). Then, the “Next Location” network determines that the next item will be stored in the third row (4). Finally, an order cue to the third row is stored in the ninth row of the “Order” network (5), and the process repeats. Here, the next encoded item will be stored in the third row of the “Item” network.

Network Operation: Recall

- To perform recall, a “Next Location” storage buffer is used by executive recall processes.
- The “Next Location” buffer stores the location of the next to-be-recalled item.
- The executive control processes can step through the “Order” buffer without accessing “Item” information.
- External order cues can also be used to access “Item” information, bypassing the “Order” network.



The above diagram shows the operation of the “Item” and “Order” networks during recall, using a “Next Location” buffer. Initially, the “Next Location” buffer indicates (1) that the next to-be-recalled item is stored in the ninth row. The phonological information in the ninth row of the “Item” network (2) is selected and sent to the LTM network (3). Meanwhile, the ninth row of the “Order” network (4) contains the information about where the next to-be-recalled item is located--in the third row. This information is retrieved and sent to the “Next Location Buffer” (5), and the process repeats.

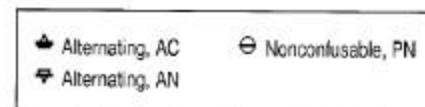
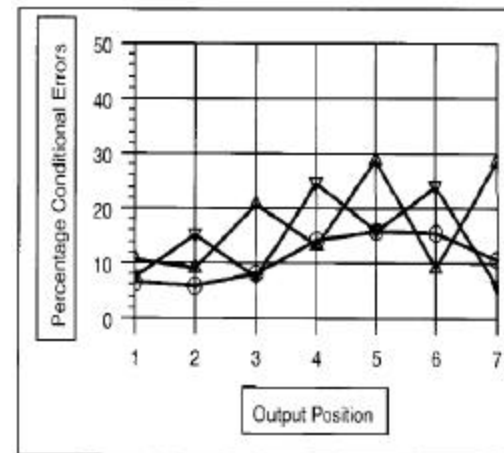
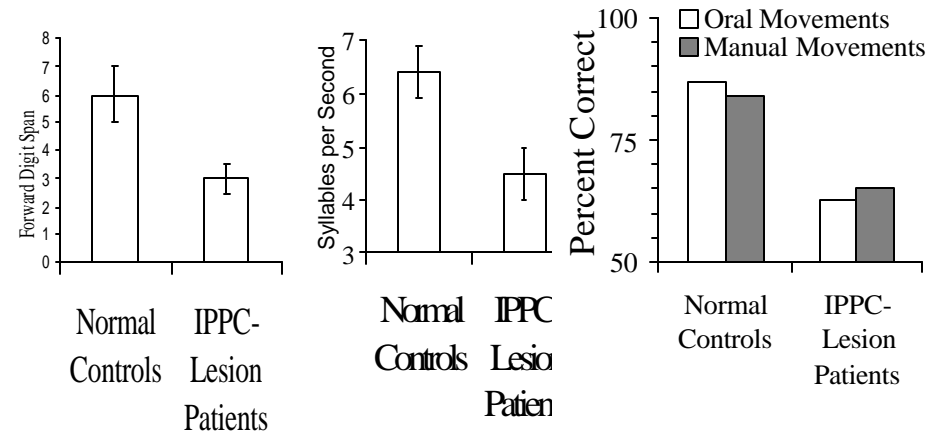
Evaluation of Proposed Model

Our model is based on several empirically testable assumptions. These assumptions include:

- Item and order information have distinct decay-time distributions.
- Item and order information are stored independently, perhaps in distinct areas of the brain.
- Item information is accessible even when order information has been lost.
- Activation levels in the network buffers increase with the amount of stored information.
- Memory for higher frequency words is better than memory for low frequency words or pseudowords.

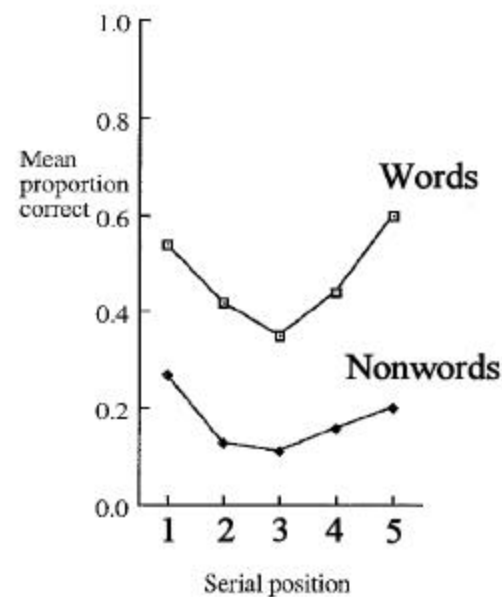
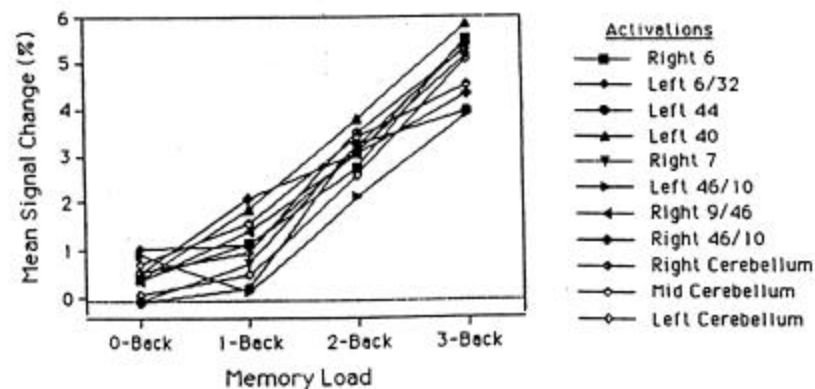
Supporting Evidence: Dissociated Memory for Item and Order Information

- Kimura and Watson (1989) found that patients with parietal damage have pervasive ordering deficits, suggesting a domain-general serial order mechanism.
- Henson et al. (1996) found that item and order information are retained quasi-independently, showing that “Order” information may be accessible even if the “Item” information is not.



Additional Supporting Evidence

- Smith, Jonides, Marshuetz, and Koeppel (1998) found that for the n-back WWM task, activation in brain regions that implement storage increases with memory load, supporting the assumptions about activation dynamics of the present model.
- Gathercole, Pickering, Hall, and Peaker (2001) found that words are recalled better than non-words, showing that well-learned stimuli are better reintegrated, as the present model predicts.



References

- Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551-581.
- Caianiello, E. R. (1961). Outline of a theory of thought-processes and thinking machines. *Journal of theoretical biology*, 11, 204-235.
- Gathercole, S. E., Pickering, S.J. Hall, and Peaker (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, 48A, 384-404.
- Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for non-words. *Journal of Memory and Language*, 35, 1-31.
- Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. New York: Wiley.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 49a, 80-115.
- Jensen, O., Idiart, M.A.P., and Lisman, J.E. (1996) Physiologically realistic formation of autoassociative memory in networks with theta/gamma oscillations: role of fast NMDA channels. *Learning and Memory*, 3, 243-256.
- Kimura, D., & Watson, N. (1989). The relation between oral movement control and speech. *Brain and Language*, 37, 565-590.
- Kohler, W. K. (1947). *Gestalt Psychology*. New York: Liveright Publishing Corporation.
- Lashley, K.S. (1930). Basic Neural mechanisms in behavior. *Psychological Review*, 37, 1-24.
- Lisman, J. and Idiart, M.A.P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, 267, 1512-1515
- Page, M., & Norris, D. (1998b). Modeling immediate serial recall with a localist implementation of the primacy model. In J. Grianger & A. M. Jacobs (Eds.) *Localist connectionist approaches to human cognition*, (pp. 227-255), Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Smith, E.E., Jonides, J., Koeppel, R.A., and Marshuetz, C. (1998) Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences*, 95, 876-882.