Proceedings of the 19th Annual Conference of the Cognitive Science Society, Mahwah, NJ:Erlbaum (1997)

The Dynamics of Prefrontal Cortico-Thalamo-Basal Ganglionic Loops and Short-Term Memory Interference Phenomena

Jack Gelfand[1], Vijay Gullapalli[1], Marcia Johnson[1], Carol Raye[1] and Jeffrey Henderson[2]
Department of Psychology[1] and Department of Computer Science[2]
Princeton University
Princeton, NJ 08544
jjg@princeton.edu

Abstract

We present computer simulations of a model of the brain mechanisms operating in short-term memory tasks that are consistent with the anatomy and physiology of prefrontal cortex and associated subcortical structures. These simulations include dynamical processes in thalamo-cortical loops which are used to generate short-term persistent responses in prefrontal cortex. We discuss this model in terms of the representation of input stimuli in cortical association areas and prefrontal short-term memory areas. We report on interference phenomena that result from the interaction of these dynamical processes and lateral projections within cortical columns. These interference phenomena can be used to elucidate the representational organization of short-term memory.

Introduction

Short term persistent response of prefrontal cortex neurons has been postulated as a mechanism for short-term memory (STM) (Fuster, 1989; Goldman-Rakic, 1994). We have previously presented models of STM based upon oscillations in prefrontal thalamo-cortical loops (Gullapalli & Gelfand, 1995; Gullapalli & Gelfand, 1997; Gullapalli, Rohde & Gelfand, 1995). In this paper we present results for the effect of lateral projections in prefrontal memory circuits on interference phenomena in STM. These mechanisms are modeled for the specific task of word processing, but the principles reviewed here are applicable to other cognitive tasks. The results of this study suggest a particular representational organization for prefrontal STM circuits.

Architecture Of Short Term Memory

The overall architecture for the postulated prefrontal STM circuits is shown in Fig. 1. The prefrontal and association cortical areas are organized as multiple groups of mutually inhibitory neurons that correspond to cortical columns (Mountcastle, 1978). These columns are indicated by groupings of neurons in the temporal association cortex and prefrontal cortex in the figure. These columns are not necessarily adjacent to each other in each cortical area but may be anatomically distributed. As discussed in the next section, we used a distributed representation over the input cortical columns with cortico-cortical interconnections between columns formed through Hebbian learning to encode stimulus words. In this representation, each column denotes a feature. Within each column, each neuron denotes a particular color or verb whose excitation is associated with that word.

Activity in the sensory/language cortical modules located in the temporal lobe association areas is transferred to the prefrontal cortical columns through direct projections. For the verbal task modeled here this would correspond to the projection from verbal association areas, such as Wernike's area, to dorsolateral prefrontal cortex (Demb et al., 1995; Snyder, Abdullaev, Posner & Raichle, 1995; Wise et al., 1991). Similarly, for a STM task with visual stimuli, this would correspond to a projection from inferior temporal cortex to prefrontal cortex (Ungerleider, Gaffan & Pelak, 1989).

In addition to receiving projections from the sensory/language cortical modules, the frontal cortex also has highly specific reciprocal projections with the thalamus, resulting in local cortico-thalamic loops. When activated, these loops can sustain activity in frontal cortex neurons (Alexander, Crutcher & DeLong, 1990; Groenewegen & Berendse, 1994; Houk, 1995; Selemon & Goldman-Rakic, 1985). These loops are activated through selective disinhibition by the basal ganglia (Chevalier & Deniau, 1990). In this model the basal ganglia function as a pattern-recognizer providing a contextual set for prefrontal cortex.

This design is similar to an architecture proposed by Wallesch and Papagano (Wallesch & Papagno, 1988) as described by Crosson (Crosson, 1992). Based on inputs from the cortical modules and the task representation, the basal ganglia selectively disinhibit the cortico-thalamic loops corresponding to the word features appropriate for the task. If the task is COLOR and the input is APPLE, for example, the column associated with color features in the prefrontal cortex would be disinhibited, and because of the feedback connections, would be allowed to oscillate. However, the specific color which would be sustained would be determined by the neuronal excitation projected from temporal association cortex to prefrontal cortex. Thus, these prefrontal neurons could serve as a working memory where

Representation of Inputs

We use a distributed representation over sensory and language cortical columnar arrays to encode stimuli. This is inspired by the functional anatomy of the cortex (Asanuma, 1975; Mountcastle, 1978; Penfield & Rasmussen, 1950). The general organization of cortical circuits is in the form of a distributed set of functionally specific regions or columns interactively involved in the representation of a given input or output. Each functionally specific region extracts from its inputs higher level information regarding a particular aspect of the task. Cortical organization in columns with reciprocal projections between columns has been observed, for example, in the

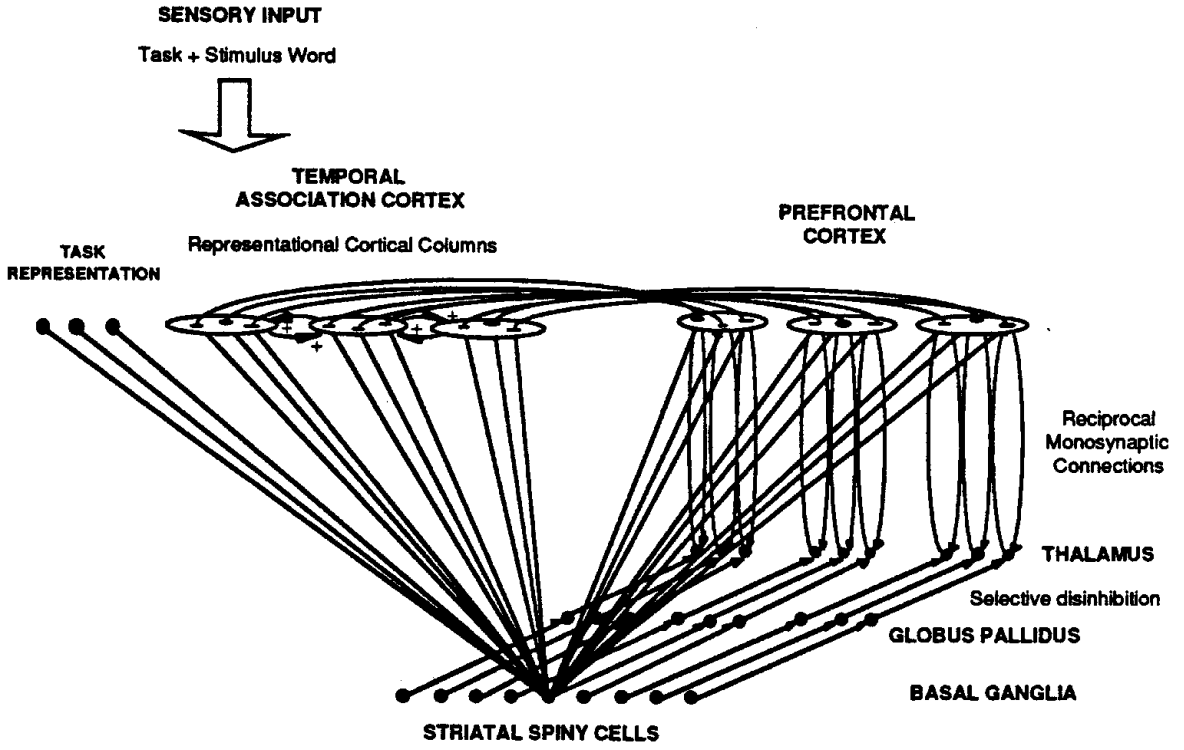

Figure 1: Architecture for association cortex and prefrontal short-term memory system. For clarity, only projections to one striatal spiny cell are shown. Similar convergent projections go to each striatal neuron. Also, the lateral inhibitory connections between neurons within each association cortex column are not shown.

task-relevant features of the stimulus are maintained for use by other cognitive or motor areas involved in the execution of the task.

Detailed descriptions of simulations using this architecture particularly with respect to the role of the basal ganglia and thalamus in task-based operation are given by Gullapalli and Gelfand (Gullapalli & Gelfand, 1995; Gullapalli & Gelfand, 1997; Gullapalli et al., 1995). In this paper we present the results of simulations to study STM interference phenomena and suggest a representational organization for the prefrontal cortical circuits involved in STM for word stimuli.

primary and secondary visual areas (Mountcastle, 1978), as well as in the motor cortex (Asanuma, 1975).

Modules in the cortical column array in our implementation, shown in Figure 2, correspond to local information processing regions of the cortex, with each module concerned with the representation of a class of feature of an input word. For example, neuronal excitation in a module might represent a color or a verb associated with the stimulus word. As a result, each word is represented as a distributed activation of a group of neurons, each encoding a feature associated with that word. In this model we employ individual neurons to encode a particular feature. However, one could use a single neuron or a distributed representation

over a group of neurons in this process with no difference in the results.

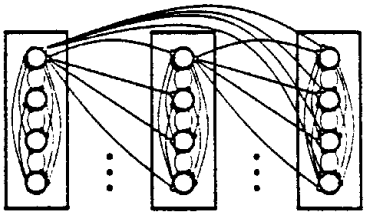

Figure 2: Block diagram of the cortical column array used in our implementation to represent words. Lines ending in open circles denote reciprocal excitatory projections between neurons in different modules, while those ending in filled circles denote inhibitory projections between neurons within a module.

For these experiments, we selected a list of 6 input words to represent. For the purposes of this simulation, the input words were represented by two cortical feature columns, one for color and one for verbs. Each contained four feature elements as shown in Table 1, e.g., black is a color and buy is a verb. The table presents the activations of the features in each column when each of the stimulus words, e.g., apple, is presented. These activations were selected to give a series of plausible responses in order to illustrate the dynamics of the system.

	apple	banana	grape	cat	dog	mouse
black	0.0	0.0	0.0	0.8	0.7	0.9
brown	0.0	0.0	0.5	0.5	0.9	0.7
red	1.0	0.0	0.3	0.0	0.0	0.0
yellow	0.0	1.0	0.3	0.0	0.0	0.0
buy	0.9	0.9	0.9	0.0	0.0	0.0
eat	0.5	0.2	0.0	0.0	0.0	0.0
fall	0.8	0.8	0.7	0.2	0.3	0.1
run	0.0	0.0	0.0	0.8	0.9	0.9

Table 1: Activations of cortical column neurons in association cortex representing stimulus words.

Proactive Inhibition and the Organization of Short-Term Memory Circuits

The organization of the brain systems simulated in this paper is supported by considerable anatomical and physiological evidence. We do not, however, have much guidance to specify the nature of the projection from posterior perceptual areas to prefrontal cortex. There is also little knowledge of the nature of representations in prefrontal cortex. We show in this paper that short-term memory interference phenomena in human subjects suggest an organization for prefrontal short-term memory circuits.

Cohen et al. propose that there is a difference in representation in posterior association cortex and prefrontal STM circuits based upon the difference in their function (Cohen, Braver & O'Reilly, 1996). They propose a distributed representation in posterior perceptual cortex similar to that described in the previous section and an independent categorical representation in prefrontal STM areas. In this paper we simulate a representational scheme in STM, which is similar to theirs, and show that it is consistent with proactive inhibition (PI) and release from proactive inhibition phenomena.

PI is a well known phenomena in which short-term memory recall is decreased due to previous related items (Wickins, 1970). In the case of words, this effect is maximum within taxonomic categories, and within the classes of stimuli such as words and numbers. That is, it is more difficult for subjects to recall items from the category toys if they have previously been asked to recall other toys in the recent past. The effect is least for words of the same part of speech or tense, and words with a similar number of syllables or phonemes. A release phenomena occurs when the subject is exposed to stimuli that are not in the same category. For example, if a subject is given a number of trials, each consisting of 3 toys, recall will diminish over trials. If a new stimulus set is presented immediately thereafter that is not in the same category, e.g., kitchen utensils, then performance will return to original levels.

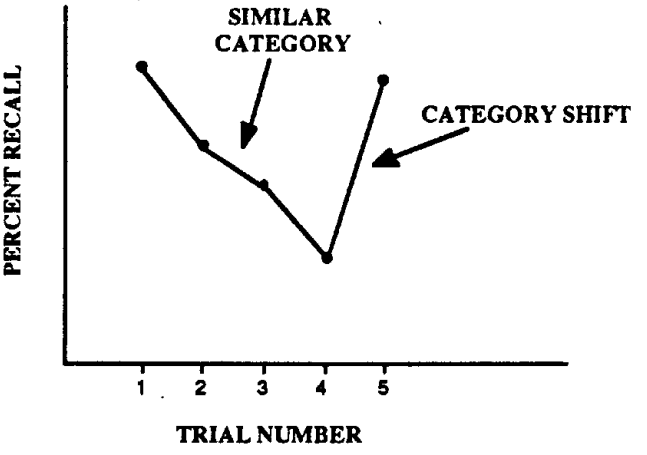

Figure 3: A schematic representation of the general phenomena of PI and release from PI

The phenomenon of PI in STM suggests that prefrontal short-term memory circuits would be made up of columns which represent individual categories of stimuli and that there are lateral inhibitory connections within those columns. These lateral inhibitory connections would provide inhibition to related concepts in a category from the residual excitation of previously presented stimuli. It also suggests that there are no lateral projections among different category columns. This independence of representation results in a restoration of performance upon exposure to new stimuli in a new class. Projections from posterior association areas would then be a pattern classifier which would transform the feature-based representation of the posterior areas to the category-based concept representation of the prefrontal short-term memory circuits.

Posterior to Anterior Projections and Representation in STM

Based upon the discussion in the last section, we have constructed an architecture for the projection from posterior association cortex to the STM area of prefrontal cortex and a representational scheme for the cortical columns in prefrontal cortex. This architecture is shown in Fig. 4. The representational scheme in the posterior association cortex is the same as that shown in Figure 2. There are two cortical columns, one each for color and verbs. The projection from the association cortex is a single layer perceptron whose input layer is the association cortex projecting to an output layer in prefrontal cortex. The single layer perceptron is a pattern classifier that classifies features in the posterior perceptual cortex and creates a representation in prefrontal cortex based upon cells representing concepts in columns organized as categories. There are two columns in prefrontal cortex organized as categories of objects, one for fruits and one for animals.

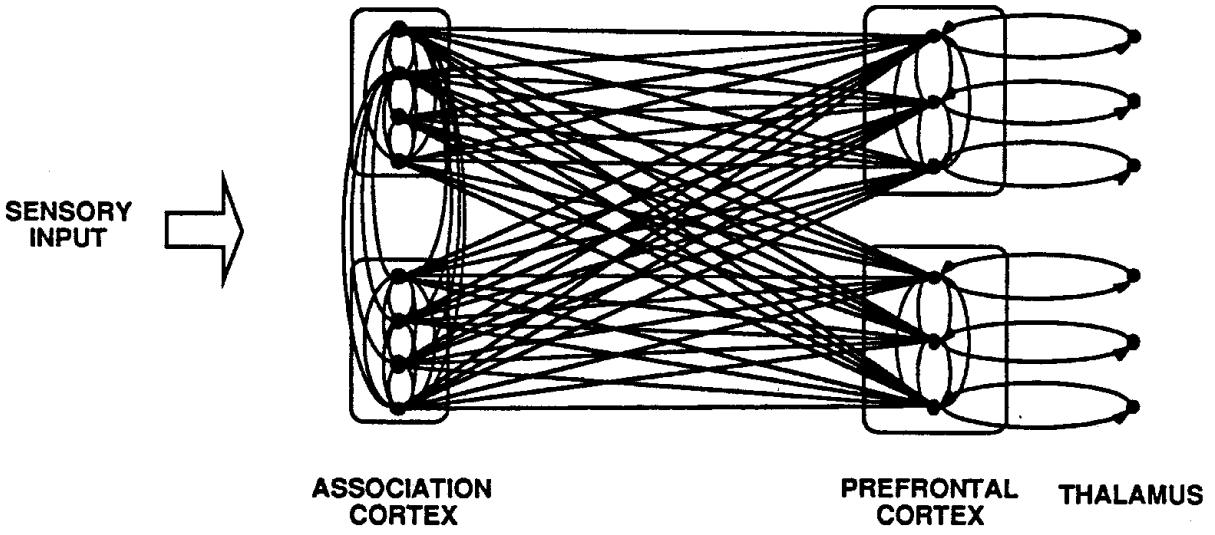

Figure 4: A schematic diagram of association cortex and prefrontal cortex and the projection between them. Lateral connections within cortical modules are inhibitory. Lateral connections between cortical columns in association cortex are excitatory. For clarity, not all intercolumn excitatory projections are shown.

Intracolumn lateral inhibitory connections and intercolumn lateral excitatory connections are shown in association cortex. Only lateral inhibitory connections are postulated to exist within columns in prefrontal cortex. No lateral excitatory projections between columns exist. As mentioned earlier, the projections from the basal ganglia and thalamus are not included in this simulation. Each neuron in prefrontal cortex has a recurrent connection with itself simulating the reciprocal connections with thalamic neurons in the full model.

Results

Persistence in Prefrontal Short-Term Memory Circuits

In this section, we present experimental results that demonstrate the properties of the model architecture. To demonstrate oscillatory behavior in the cortico-thalamic loops, we presented three sequential stimuli to the input circuit. The graph in Fig. 5 shows the activity of neurons in association cortex and prefrontal cortex. The stimulus interval for each sequential stimulus is shown above the graph denoted by the first letter of banana, apple and grape, respectively. In each case we note that the activity of neurons representing the component features of each stimulus is excited in association cortex and decays rapidly after the stimulus is terminated. The neurons in prefrontal STM memory cortex, however, continue to fire after the stimulus is removed due to thalamo-cortical oscillations. The oscillations eventually decay because of the leaky neurons used in the circuit. This captures the dynamics of thalamo-cortical loops suggested by several researchers (Alexander et al., 1990; Chevalier & Deniau, 1990; Fuster & Alexander, 1973; Goldman-Rakic & Friedman, 1991; Wang, Rinzel & Rogawski, 1991).

Simulation of Proactive Inhibition in Short-Term Memory

PI and release from PI in STM was simulated by presenting three sequential stimuli in the fruit category followed immediately by three sequential stimuli in the animal category. In each case the new stimulus was presented when the level of activation of the previous stimulus was at about 25% of its peak level. The level of activation for the neuron representing each stimulus is shown in Fig. 6 compared to the level of activation of each stimulus neuron if the stimulus was presented alone. The triangular points denote activation that would result for

separate trials where there is no intertrial interference. The circles denote activation for trials presented in succession where residual activity of previous stimuli through lateral inhibitory connections causes a diminished response for subsequent stimuli. When there is a shift in the category of stimulus from fruits to animals, the activity of the prefrontal STM neurons recovers to a level which is equal to the activity of that neuron if the stimulus was presented alone. This decrement and recovery in activity levels in STM would result in the shifts in recall performance shown in Fig. 3. We note a slight decrement of activity for individual exposures to the animal stimuli in the cat, dog mouse ordering shown in Fig. 6. We found that this is due to lowered activity in the association cortex because many shared features in these stimuli result in greater total lateral inhibition. This is due to the simple nature of lateral projections in the model. We found that the effects reported here are not dependent upon the initial choice of stimuli or their relationships.

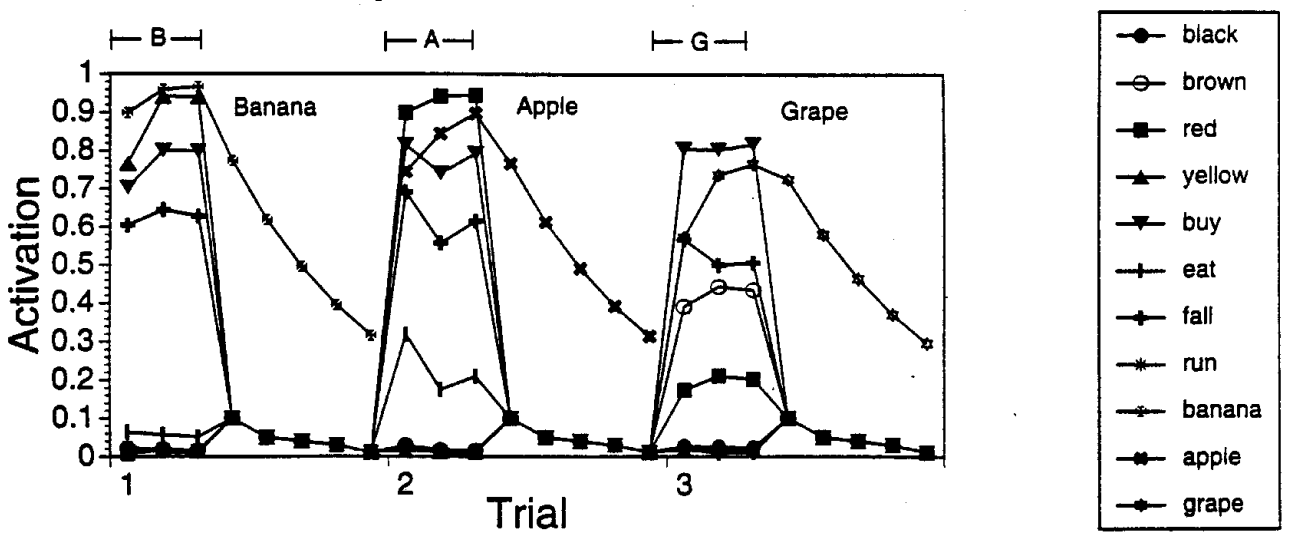

Figure 5. Examples of the activation of association cortex and prefrontal cortex neurons when various stimuli are presented. The activity of the neurons labeled banana, apple and grape are in prefrontal cortex. The others are in association cortex.

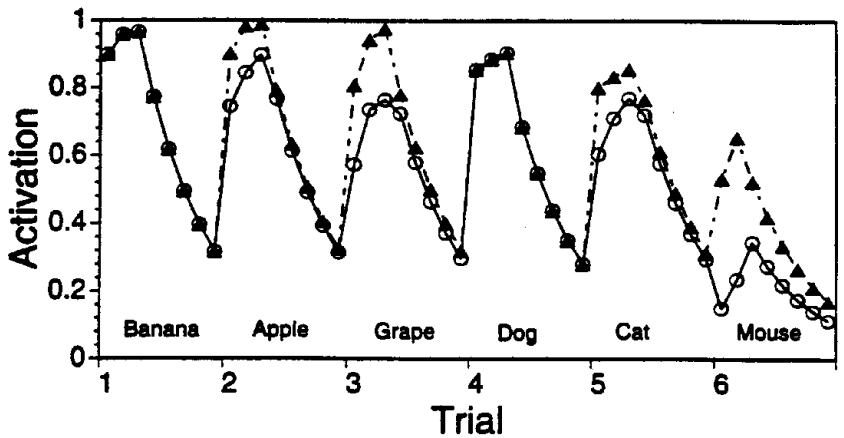

Figure 6: A graphical representation of the simulation of proactive inhibition and release from proactive inhibition. The activation of neurons in prefrontal cortex is plotted as a function of time during the presentation of 6 stimuli as noted on the graph. The triangular points denote activation that would result for separate trials where there is no intertrial interference. The circles denote activation for trials presented in succession where residual activity of previous stimuli through lateral inhibitory connections causes a diminished response for subsequent stimuli.

Discussion

Though the phenomenon of PI and release from PI could be simulated quite readily with the postulated representational organization of STM, it is possible that these phenomena could also result from a representation in STM based upon features, as in the organization of association cortex in the model. PI and release from PI could be simply due to the fact that objects in the same category share more features than objects in other categories. Future simulations in combination with studies of subjects' performance will be directed at this issue.

The model used in the simulations reported here is quite simple. The features of the model highlight the fact that

representational issues can be elucidated by simulations of interference phenomena. We did not include the role of the basal ganglia in controlling thalamo-cortical oscillations by disinhibiting specific thalamic neurons. This would lead to an attentional phenomenon which may interact strongly with the interference phenomena we observed.

Acknowledgments

This work was supported by a grant from the James S. McDonnell Foundation and NIA Grant AG09253.

References

Alexander, G., Crutcher, M., & DeLong, M. (1990). Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In H. Uylings, C. Van Eden, J. De Bruin, M. Corner, & M. Feenstra (Eds.), *The Prefrontal Cortex: Its Structure, Function and Pathology* (pp. 119-146). Amsterdam: Elsevier.

Asanuma, H. (1975). Recent developments in the study of the columnar arrangements of neurons in the motor cortex. *Physiological Review*, 55, 143-156.

Chevalier, G., & Deniau, J. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in Neuroscience*, 13, 277-280.

Cohen, J., Braver, T., & O'Reilly, R. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: Recent developments and current challenges. *Phil. Trans. R. Soc. London. B*, 351, 1515-1527.

Crosson, B. (1992). Chapter 4, Theories of subcortical structures in language, Subcortical Functions in Language and Memory (pp. 111-144). New York: Guilford Press.

Demb, J., Desmond, J., Wagner, A., Vaidya, C., Glover, G., & Gabrieli, J. (1995). Semantic encoding and retrieval in the left prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, 15, 5870-5878.

Fuster, J. (1989). *The Prefrontal Cortex*. (2nd ed.). New York: Raven Press.

Fuster, J., & Alexander, G. (1973). Firing changes in cells of the neucleus medialis dorsalis associated with the delayed response behavior. *Brain Research*, 61, 79-81.

Goldman-Rakic, P. (1994). The issue of memory in the study of the prefrontal cortex. In A. Thierry, J. Glowinski, P. Goldman-Rakic, & Y. Christen (Eds.), *Motor and Cognitive Functions of the Prefrontal Cortex* Berlin: Springer-Verlag.

Goldman-Rakic, P., & Friedman, H. (1991). The circuitry of working memory revealed by anatomy and metabolic imaging. In H. Levin , H. Eisenberg, & A. Benton (Eds.), *Frontal Lobe Function and Dysfunction* (pp. 72-91). New York: Oxford University Press.

Groenewegen, H., & Berendse, H. (1994). Anotomical relationships between the prefrontal cortex and basal ganglia in the rat. In A. Thierry, J. Glowinski, P. Goldman-Rakic, & Y. Christen (Eds.), *Motor and Cognitive Functions of the Prefrontal Cortex*. Berlin: Springer-Verlag.

Gullapalli, V., & Gelfand, J. (1995). A model of the dynamics of prefrontal cortico-thalamo-basal ganglionic loops in verbal response selection tasks. In J. Grafman, K. Holyoak, & F. Boller (Eds.), *Structure and Functions of the Human Prefrontal Cortex* (pp. 375-380). New York: New York Academy of Sciences.

Gullapalli, V., & Gelfand, J. (in press). Neural modeling of learning in verbal response selection tasks. In J. Donahoe & V. Dorsel (Eds.), *Neural Network Models of Cognition: Biobehavioral Foundations*. Amsterdam: Elsevier Science.

Gullapalli, V., Rohde, D., & Gelfand, J. (1995). A model of a dual-pathway system for practice-related learning in verbal association tasks. In J. Moore & J. Lehman (Eds.), *Proceedings of the 17th Annual Cognitive Science Conference* (pp. 136-141). Mahwah, NJ: Erlbaum.

Houk, J. (1995). Information processing in modular circuits linking basal ganglia and cerebral cortex. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 3-9). Cambridge: MIT Press.

Mountcastle, V. (1978). An organizing principle for cerebral function: the unit module and distributed function. In G. Edelman & V. Mountcastle (Eds.), *The Mindful Brain* Cambridge: MIT Press.

Penfield, W., & Rasmussen, T. (1950). *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York: Macmillan.

Selemon, L., & Goldman-Rakic, P. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *Journal of Neuroscience*, 5, 776-794.

Snyder, A., Abdullaev, Y., Posner, M., & Raichle, M. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proc. Nat. Acad. Sci.*, 92(5), 1689-1693.

Ungerleider, L., Gaffan, D., & Pelak, V. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Experimental Brain Research*, 76, 473-484.

Wallesch, C.-W., & Papagno, C. (1988). Subcortical aphasia. In F. Rose, R. Whurr, & M. Wyke (Eds.), *Aphasia* (pp. 256-287). London: Whurr Publishers.

Wang, X., Rinzel, J., & Rogawski, M. (1991). A model of the {T}-type calcium current and the low-threshold spike in thalamic neurons. *Journal of Neurophysiology*, 66, 839-850.

Wickins, D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77, 1-15.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, 114, 1803-1817.