A Model That Accounts for Activity in Primate Frontal Cortex during a Delayed Matching-to-Sample Task

Sohie Lee Moody,[1,2] Steven P. Wise,[2] Giuseppe di Pellegrino,[2] and David Zipser[1]

[1]Cognitive Science Department, University of California, San Diego, La Jolla, California 92093-0515, and [2]Laboratory of Systems Neuroscience, National Institute of Mental Health, Poolesville, Maryland 20837

A fully recurrent neural network model was optimized to perform a spatial delayed matching-to-sample task (DMS). In DMS, a stimulus is presented at a sample location, and a match is reported when a subsequent stimulus appears at that location. Stimuli elsewhere are ignored. Computationally, a DMS system could consist of memory and comparison components. The model, although not constrained to do so, worked by using two corresponding classes of neurons in the hidden layer: storage and comparator units. Storage units form a dynamical system with one fixed point attractor for each sample location. Comparator units constitute a system receiving input from these storage units as well as from current input stimuli. Both unit types were tuned directionally. These two sources of information combine to create unique patterns of activity that determine whether a match has occurred. In networks with abundant hidden units, the storage and comparator functions were distributed so that individual units took part in both. We compared the model with single-neuron recordings from premotor (PM) and prefrontal (PF) cortex. As shown previously, many PM and PF neurons behaved like storage units. In addition, both regions contain neurons that behave like the comparator units of the model and appear to have dual functionality similar to that observed in the model units. No neuron in either area had properties identical to those of the match output neuron of the model. However, four PF neurons and one PM neuron resembled the output signal more closely than any of the hidden units of the model.

Key words: model; attractor; prefrontal; premotor; neural network; matching-to-sample; comparator

Delayed matching-to-sample tasks (DMS) have been important in studying the neural basis of short-term memory. In spatial versions of these tasks, a stimulus is presented at one location as a *sample*. Later, stimuli (termed *current* stimuli) are presented at either the location of the sample (*matches*) or at different locations (*distractors*). Matches are to be reported. DMS requires two major functional components: (1) memory, buffered against distractors, to hold the sample; and (2) comparison of current and stored stimuli to detect matches. From a computational viewpoint, these two functions can be implemented using an attractor neural network, which has stable points to store the sample value, and a filter to detect matches.

The first of the two functional components of a DMS network have been studied in both theoretical work and neural modeling. Short-term information storage can be accounted for by fixed point attractor dynamics of neural activity in recurrently connected networks (Cowan, 1972; Zipser, 1991; Amit and Brunel, 1995; Zhang, 1996). Furthermore, neuronal activity consistent with this mode of information storage has been reported widely, e.g., in prefrontal (Fuster and Alexander, 1971), supplementary motor (Tanji et al., 1980), premotor (Weinrich and Wise, 1982), primary motor (Evarts and Tanji, 1976), somatosensory (Zhou and Fuster, 1996), posterior parietal (Crammond and Kalaska, 1989), auditory (Vaadia et al., 1982), and visual (Mikami and Kubota, 1980) cortex. Buffering of stored information against distractor events has also been reported. It appears to occur in frontal cortex (di Pellegrino and Wise, 1993a,b; Miller et al., 1996) but not in temporal (Miller et al., 1996) or posterior parietal areas (Steinmetz et al., 1994; Steinmetz and Constantinidis, 1995). Recent neuroimaging studies have yielded corresponding results in human subjects regarding storage and buffering (Cohen et al., 1997; Courtney et al., 1997). However, the second component of a DMS network, involving comparison and matching, has not been addressed.

To investigate how the brain may implement both major components of a DMS network, we trained a recurrent neural network to perform spatial DMS. This approach, called *neural systems identification*, has been shown to generate realistic models (Zipser and Anderson, 1988). Analysis of the trained network revealed that the underlying computational solution indeed uses an attractor network to store sample information, as well as a set of comparator neurons, the combined action of which approximates a match filter. The implementation of the model involved specific features that are not obvious consequences of the basic computational mechanism, such as directional tuning of both the attractor network and the comparator neurons, together with a considerable degree of distributed function. When frontal cortex activity was compared with the model, directionally tuned units resembling comparator and attractor network neurons were found, along with evidence for distributed function. These results suggest that a computational solution similar to that in our model may be implemented in a network that includes frontal cortex neurons.

MATERIALS AND METHODS

Modeling methods. The model described in this paper was generated using a technique called neural systems identification (NSI) (Zipser, 1992). As this technique was applied here, an initially randomly

Received May 13, 1997; revised Oct. 10, 1997; accepted Oct 15, 1997.

Correspondence should be addressed to Dr. Sohie L. Moody, Laboratory of Systems Neuroscience, National Institute of Mental Health, P.O. Box 608, Poolesville MD 20837.

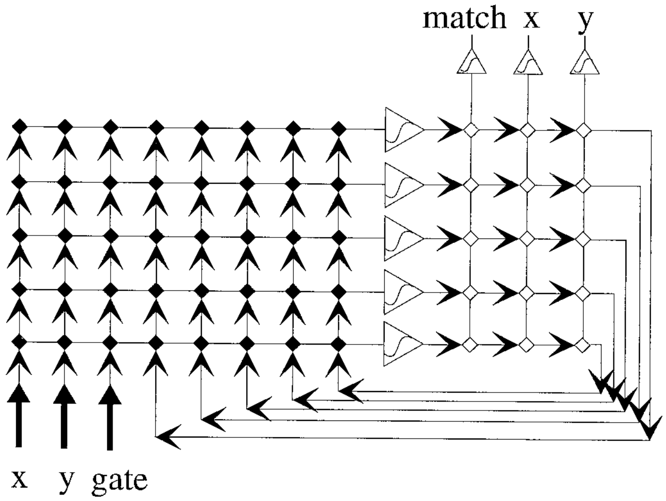

Figure 1. Diagram of the recurrent network model. Each *large, sigmoid-containing triangle* represents a soma or a model unit, which receives external input and feedback from other network units. The activation of a model unit is determined by performing a nonlinear operation on the weighted sum of its inputs (see Eq. 1). Inputs are shown in the *bottom left corner* and labeled *x*, *y*, and *gate*, respectively. For clarity, only five model units are shown here; actual network models contained between 9 and 50 of these units. The *black diamonds* represent the w_{ij}; *white diamonds* represent output unit weights. The three *small, sigmoid-containing triangles* labeled *match*, *x*, and *y* represent the output units. The activation of an output unit is determined by a nonlinear operation on the weighted sum of the hidden unit activity.

weighted, fully recurrent network of simple model neurons was trained to emulate the input–output behavior conjectured to occur in frontal cortex during the DMS task. Neural systems identification differs from most conventional modeling techniques in that the model is generated from a general conjecture about the function of a brain area, rather than being constructed from detailed knowledge or hypotheses about the internal structure of that area. This approach has the advantage that identification models often generate mechanisms that otherwise would not be considered. Moreover, NSI spawns model neurons with response properties closely resembling those of cortical neurons. Identification models, in their current form, have the disadvantage that they are limited in the amount of realistic detail they provide. They generally give information only about the average activity of component neurons.

Neural systems identification generates models by adjusting the synaptic weights recurrently connecting the internal, or hidden, units in the network so that they function to perform the desired input–output behavior. The process of evolving a set of synaptic weights that minimizes the error produced by the network is called optimization or "training." The results are typically independent of the exact optimization procedure used, because the set of possible final configurations of the model depends on the task rather than on how the optimization was done. For the model described here we used a gradient descent, error correction optimization algorithm for recurrent networks, called back-propagation through time (Williams and Zipser, 1995). Because the functions of the hidden units develop during the model generation process and are not specified at the beginning of training, *post hoc* analysis of the mechanisms of the model for solving the input output transform may yield insight into how biological networks may solve similar problems.

The model network is shown in Figure 1. It consists of a set of fully interconnected model neurons, simulated with logistic functions. The model is implemented by the system of nonlinear difference equations shown below:

$$y_i(t+1) = f\left(\sum_j w_{ij} y_j(t) + \sum_k v_{ik} p_k(t) \right) \qquad (1)$$

where $y_j(t)$ are the output values of all other model neurons at time t; w_{ij} are the weight values connecting model neuron i to model neuron j; $p_k(t)$ are the external inputs at time t; v_{ik} are the weights connecting model neuron i to input line k; and $f(x)$ is the logistic function $f(x) = 1/(1 + e{-}^{x})$.

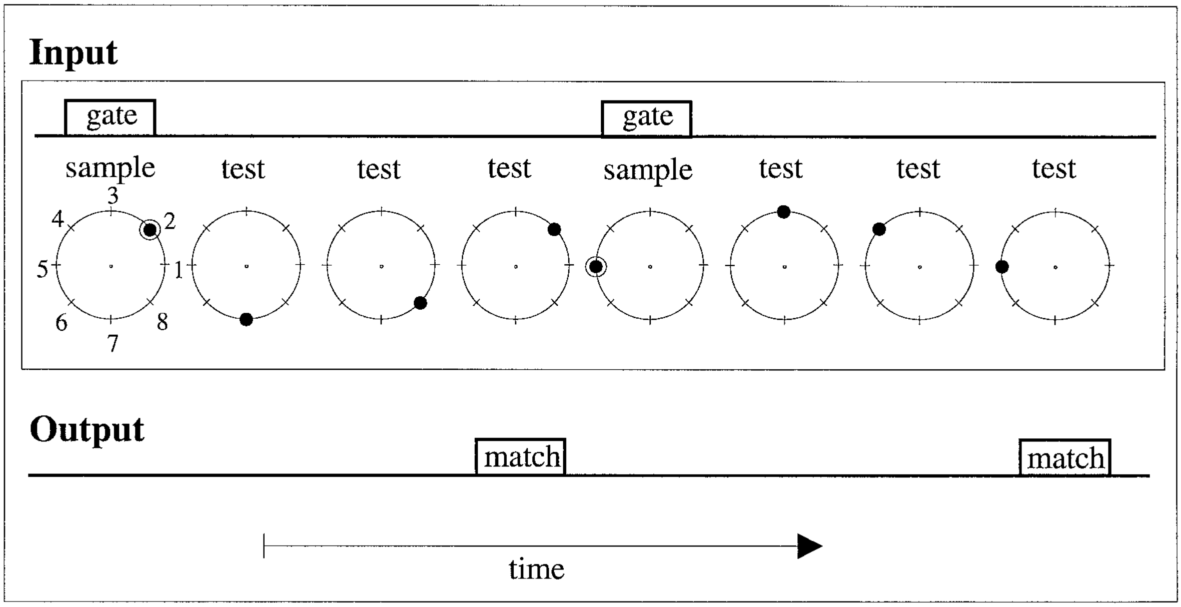

Figure 2. Schematic of training algorithm used for delayed match-to-sample network. The *row* of *circles* represents the input coordinate stimuli, and the *bottom line* represents the output match response. The *top line* represents the gate signal. There are eight possible input coordinates, represented by the *small lines* on the *circles*. A *solid circle* indicates an input stimulus pattern; an *outlined circle* indicates a gated stimulus pattern. The output match unit indicates when the current input matches the last gated input. The output also carries a copy of the last gated location, which is not shown. Each gated sample input is followed by a pseudorandom number of distractors. The number of distractors, n, decreases exponentially as n increases and is determined as the probability, $p(n)$, that the number of test patterns is $n = k^{(n-1)}(1 - - k)$, where k was set to 0.35. In addition, there was an interstimulus interval of either two or three (*0,0*) inputs separating each new input stimuli.

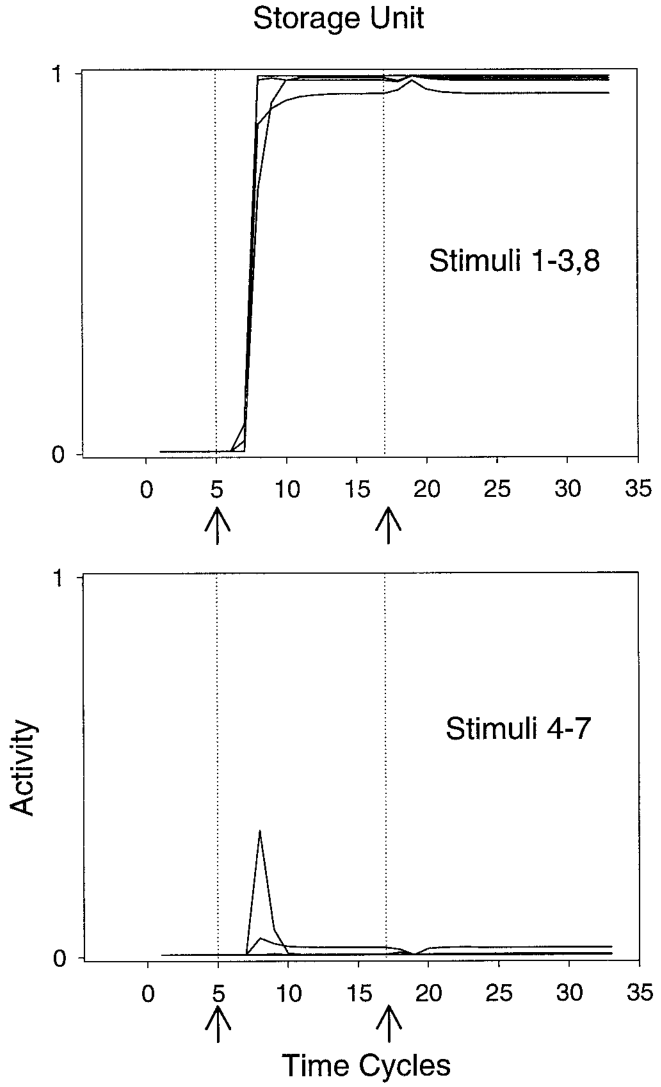

Figure 3. Typical temporal behavior of a the storage unit of a model, shown as unit activity plotted as a function of time. At t = 5, a sample stimulus pattern is gated in, and at t = 17, the same stimulus is presented again. Both events are marked with *arrows*, and *dotted vertical lines* demarcate the delay period. The *top plot* shows the sustained response for a subset of input patterns (sample stimuli at locations 8 and 1–3); the *bottom plot* shows the sustained response for the remaining inputs (sample stimuli at locations 4–7). In this and most subsequent figures, model unit activity levels are normalized to the range [0.0–1.0].

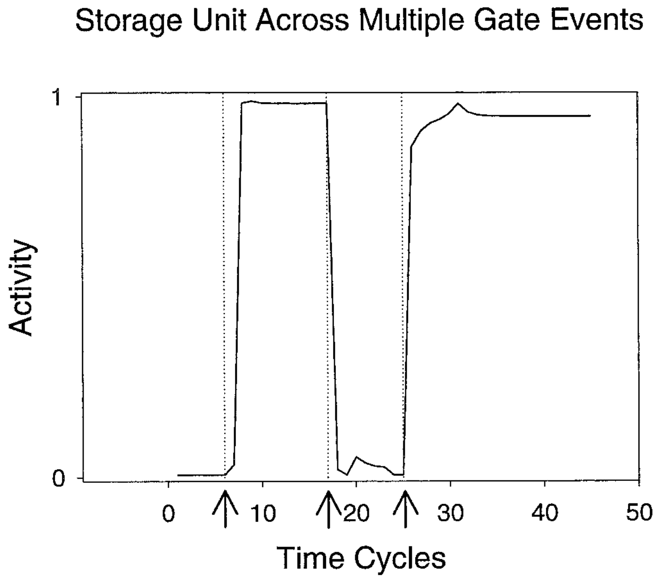

Figure 4. Temporal behavior of a storage unit across multiple gated sample input patterns. Each *arrow* along the x-axis indicates gating in of a new sample pattern. At t = 6 and 25 a "preferred" sample location was gated in; at t = 17 a nonpreferred sample location was presented.

In our model, there are three input lines: two for the (x,y) coordinate of the stimulus and one called the "gate." The gate takes a value of 1 when new information is to be stored and is 0 otherwise, including when distractor and matching stimuli are applied to the stimulus position inputs. These three inputs, then, provide the information that *must* be available solve the DMS problem. Other input formats, e.g., retinotopic coordinates, could have been chosen, but the exact format used here was chosen for simplicity. Whereas the information content and dimensionality of NSI model inputs are critical, the precise *format* of inputs generally has only a second-order quantitative effect and does not change the major functional properties of the model (Moody and Zipser, 1998; Sanger, 1994). The input lines connect to all model neurons except the output units. The output of the network consists of a set of three units that receive input from all of the hidden units but do not feed back to them. In addition to the match output unit there are two additional output units that indicate the coordinates of the stored stimulus position. These output units acted as an external copy of the location to be remembered, thereby facilitating the match process. The units were found to be required for efficient training but were not necessary for the basic properties of the hidden units or the overall performance of the model.

To train the model, randomized sequences of stimulus positions and gate signals were provided at appropriate times. The output was trained to indicate when a match occurs, as well as the location of the previously stored stimulus. Figure 2 provides details of the training paradigm. Gate signals were separated by intertrial intervals in which the stimulus position was set to zero, i.e., at the center of the circle of stimuli positions. Models were generated by training networks for about 10 million network time steps. On average, three distractors were shown before presentation of an input that matched the stored information. The average total duration from sample to match was 15 time steps. Starting from random weights, networks of nine or more units always learned to do the task correctly.

Neurophysiological methods. We compared the response properties of hidden units in our model with those of frontal cortex neurons recorded in previously reported experiments (di Pellegrino and Wise, 1993a,b). The experimental details can be found in the previous reports but are briefly summarized here. A rhesus monkey was trained in a manner roughly analogous to that of the model described above. In the experimental condition used for the present analysis (termed the "compatible condition" by di Pellegrino and Wise, 1993a,b), each trial began with the monkey centering a two-joint manipulandum beneath a light-emitting diode (LED) that was in the center of a circular array of eight LEDs. The monkey fixated a central LED, and later, one of the eight peripheral LEDs was illuminated for 500 msec as the sample stimulus. After a delay of either 550 or 750 msec, a current stimulus was presented for 100 msec. If it was one of the seven LEDs other than the sample, no response was required, and any motor response terminated the trial. If the current stimulus matched the location of the sample, then the monkey had to move the manipulandum to that location within 650 msec to report that a match had occurred. (This positional motor response is analogous to the analog information contained in the x and y output units of the network.) If performed accurately, juice reinforcement was delivered; otherwise the trial terminated. The present analysis focuses on 68 neurons recorded from the dorsal premotor cortex (PM), a part of the frontal cortex thought to be involved in visually guided movement. In addition, 37 neurons from part of the prefrontal cortex (PF) were also analyzed.

Analytical methods. To quantify the degree and depth of direction tuning, we developed four indices, computed for a given unit across all eight directions. Similar types of indices have been used in the past, often based on fitting the data to a cosine function (Georgopoulos et al., 1982, 1988). However, inspection of the neurophysiological data indicated that

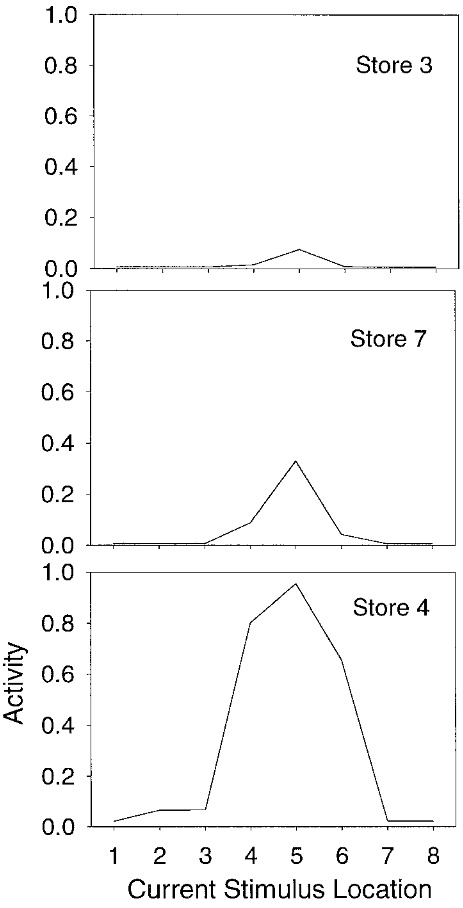

Figure 5. Model comparator unit for three different stored values. The response to current stimuli, matches, and distractors at locations 1–8 when the sample has previously appeared at location 3 is shown in the *top plot*. The *middle* and *bottom plots* show the response to current stimuli at the same locations for sample stimulus locations 7 and 4, respectively.

regression to a cosine function would not yield many good fits. Accordingly, we developed indices that did not depend on sinusoidal directional tuning. A selectivity index (s_i), depth-of-tuning index (d_i), activity level index (a_i), and modulation index (m_i) are defined as follows:

$$s_i = \frac{k - \left(\frac{\sum_{n=1,k} i_n}{i_{\max}} \right)}{k-1} \tag{2}$$

$$d_i = \frac{i_{\max} - i_{\min}}{i_{\max}} \tag{3}$$

$$a_i = i_{\max} \tag{4}$$

$$m_i = \max \mid \vec{i}_{\text{match}} - \vec{i}_{\text{delay}} \mid \tag{5}$$

where k is the number of directions; $\vec{i}_{\text{event}}$ denotes the activity level during a given event period (such as sample, delay, or match) for a given cell across eight directions; and $i_{\min}$ and $i_{\max}$ are, respectively, the minimum and maximum responses of the i^{th} neuron across the eight different directions.

RESULTS

Mechanism of the model

To determine the mechanism of the model, several analytical approaches were used, based on the following properties of the model neurons: (1) the activity pattern within a simulated trial; (2) the steady-state activity pattern during delay period; (3) the consequences of single-unit lesions; and (4) the synaptic weight connections.

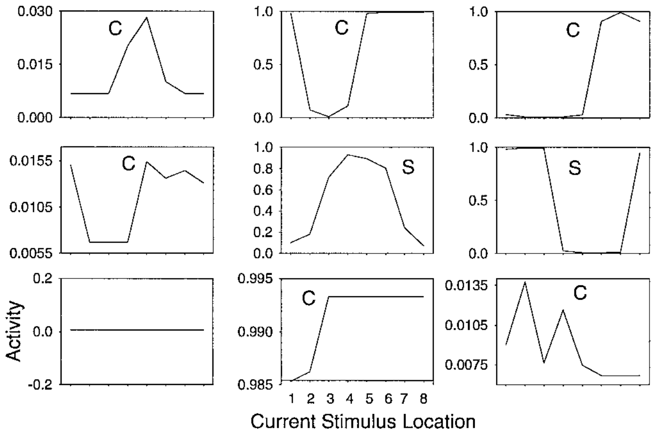

Figure 6. Steady-state values of the network, for a minimal network, shown as activation profiles of the nine model neurons across eight input patterns. The activation shown is after the network has settled to a steady state. *C*, Comparator cell; *S*, storage unit. Note how one unit (*bottom left corner*) does not distinguish among the different input patterns.

An important parameter in NSI models is the number of hidden units used. We found significant differences between the behavior of hidden units in minimal versus larger models. The hidden units in minimal size networks were of relatively "pure" functional types, with each unit devoted to one of the main information-processing components of the task. In larger models, only a fraction of the units were of pure functional types. Most units seemed to be participating in several of the necessary information-processing components, with varying degrees of participation in each. This distinction between minimal and larger networks allowed us to use the operational mechanisms of minimal networks, which were fairly transparent, to gain insight into the potential mechanism of more complex networks, which were more difficult to analyze. It also allowed us to gain further insight into the neurophysiological data, because the biological units more closely resembled the hidden units in large networks than in minimal ones.

At the start of a typical simulated behavioral trial, the center coordinate (*0,0*) was gated into the model network, corresponding to the starting position in a reaching task. After an intertrial interval, a stimulus location was gated in. This was followed by zero to three distractor stimuli and finally by the match stimulus. Analysis of the data from these simulations revealed two distinct kinds of hidden units: storage and comparator units.

Storage units maintain a sustained level of activation that is determined by the last input loaded simultaneously with a gate pulse. A new gate pulse is needed to cause a change in the activity of a storage unit (see Fig. 4). Storage units are directionally tuned and quite insensitive to postgate events, such as distractors and matches (Figs. 3, 4).

Comparator units receive input from both the current stimulus and the stored sample. They can respond to distractors as well as to match stimuli, and their responses are directionally tuned to the location of the current input stimulus. In the majority of comparator units, amplitudes of response to current stimuli are modulated by the value held by the storage units. This modulation includes both potentiation and suppression. An example of how a

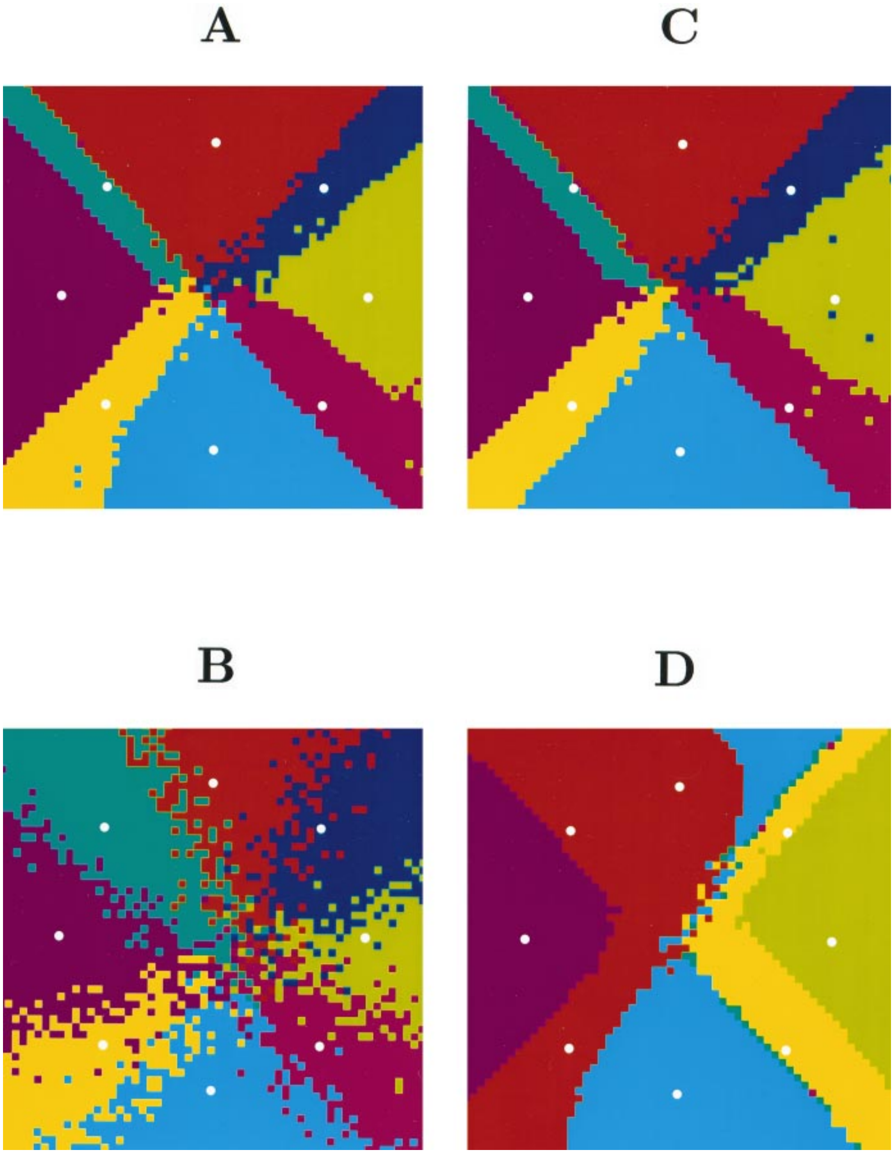

Figure 7. Spatial extent of the basins of attraction for model networks and the effects of lesions on attractor basins. *A*, Minimal, nine-unit network. *B*, Fifty-unit network. *C*, Nine-unit network with a lesioned comparator unit. *D*, Nine-unit network with a lesioned storage unit. Steady network state activation is plotted as a function of the location of the gated-in (x,y) coordinate sample stimulus across a matrix of 57 × 57 (x and y ranged from $[-0.2{:}1.2]$, with a step of 0.025). The network settled for 25 activation sweeps after each gated coordinate position. The various *colors* serve to distinguish one attractor basin from another; there is no correlation between hue and relative distances among attractor basins. The *small white circles* indicate the eight training target locations. Note that, because of storage unit activity, the basins of attraction remain stable with a comparator lesion (compare *A*, *C*). Note the serious disruption of several basins of attraction with a storage unit lesion (compare *A*, *D*), including one large basin that now subsumes three of the eight locations in the sample stimulus training set.

comparator unit responds to all current stimuli for various stored samples is shown in Figure 5. Note that the response is directionally tuned, and that the shape and peak of the tuning curve remain fixed for different stored samples. It is the magnitude of the tuning curve that is dependent on the current stored sample. This relatively fixed shape of the tuning curve for current stimuli allows us to incorporate both the current stimulus and the stored sample by describing the output of a comparator unit as approximately their product. Thus, the output of the network can be represented, to a first approximation, as:

$$m_{jk} = \sum_{i} w_i a_{ij} a_{ik} \tag{6}$$

where m_{jk} is the output of the match unit, w_i is the weight connecting the ith comparator unit to the output, and $a_{ij}a_{ik}$ is the product of the response to the jth current stimulus and the kth stored sample of the ith comparator unit. This equation has the general form of a match filter that signals a match when the sum m_{jk} exceeds a preset threshold (Oppenheim et al., 1996). In the network model, this threshold is implemented by the sigmoidal activation function of the output unit, thereby enabling the network output to detect the eight match conditions of the 64 possible combinations of current stimulus and stored sample. A strong indication that the output match unit uses the responses of comparator units to make its decision can be seen in the large magnitude of the weights that connect them to the match output unit. The analog of Figure 5, using experimental data, is shown in Figure 12. It is quite noisy but suggests that a similar computational mechanism may be used by the brain. It is interesting that no units were found in the recurrent layer that carry the match decision signal. No true match units were found in the cortical areas studied either, as will be discussed later.

When the number of hidden units is increased from the minimum needed, a significant difference appears in the properties of many of the hidden units. There are still some relatively pure functional types, but most hidden units in larger networks are of complex, mixed types. These mixed types have activity profiles that seem to combine storage and comparison functions to various degrees. Analysis of the operations of the model, by examining the effects of removing single units from the network, demonstrate that these composite functional units play an essential role in the DMS task.

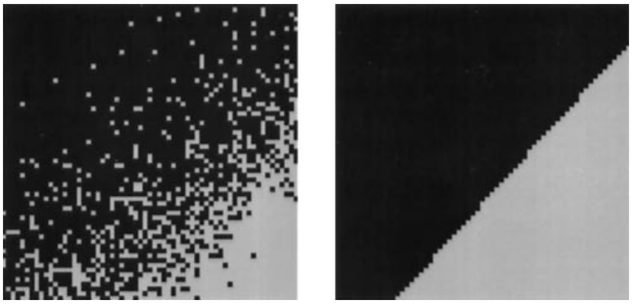

Figure 8. Higher-resolution map of a boundary between two basins of attraction. These plots show an x range of [0.670–0.794], a y range of [0.540–0.664], and a step size of 0.002 (vs 0.025 in Fig. 7). *Left*, Each hidden unit was initialized with a different random value. *Right*, Each hidden unit was initialized with the same value.

Additional information about the model was obtained by studying its steady states. All instances of the model settled to steady states; that is, the set of model neurons evolved to a state that was changeless over time. The steady states were found by gating in a stimulus and then returning the input lines to zero until the activity of all units became fixed over time. The effect of stored information on the steady-state values of all the units in a minimal size network is shown in Figure 6. This instance of the model has two storage units (labeled *S*), six comparator units (labeled *C*), and one unit that is directionally tuned to gate events but is not modulated by stored value. The two storage units are directionally tuned to the stored information, but one has eight distinct values over direction, and the other has just two values, high and low. Within the set of six storage-modulated comparator units, all but one of the steady-state tuning curves are approximately unimodal. Note that the steady-state tuning curves for the six tuned comparator units are all different. These differences play a central role in enabling the comparator units to provide unique information about all combinations of input and stored information. This information is then used by the match output unit for recognizing matching stimuli.

The stable states of our model are fixed point attractors. Fixed point attractors have the property that if the system is given a small perturbation from its steady state it will return to that state. The region of state space from which the system will evolve to a specified attractor is called a basin of attraction. In our model, the remembered stimuli are stored in basins of attraction. There are eight attractors, one for each of the eight possible stimuli in the training set. It is often hard to visualize basins of attraction, because they are in a space with the same number of dimensions as the number of units in the network. However, the attractor structure of our model networks is relatively simple and could be visualized in the two-dimensional space that contains the eight input stimuli. To plot the basins of attraction, we systematically gated in a square grid of 57 × 57 (3249) points covering the two-dimensional input space and including the eight training stimuli locations. To maximize the number of attractors found, all model neurons were set to randomly chosen values before presentation of a new grid location. After gating in a new location, the input lines were held at zero until the network settled to a stable state. Two stable states were considered the same if each model neuron in one stable state had the same activity level as the corresponding neuron in the second stable state. This procedure produced only eight stable states for all 3249 points (Fig. 7). Figure 7, *A* and *B*, shows these eight stable states for a nine unit and a 50 unit network model. Each of the eight stimuli locations on which the network was trained was surrounded by a large basin of attraction. These eight basins of attraction accounted for all of the input space, and there were no other attractors in the network.

The robustness of the attractors was examined by studying the effect of initial network starting position on the resulting attractors. The initial state of the network refers to the collective set of initial activity level of each model neuron. The overall structure of these attractor basins was not affected by the initial state of the network; only the basin boundaries were affected. If the network always started from the same initial state, a smooth border evolved between basins (Fig. 8 *right*), whereas if the network started from a randomly selected initial state each time, then a noisy boundary resulted (Fig. 8, *left*). Thus, the general structure of the attractor basins are robust, which implies that the stored information will be buffered against noise and irrelevant neural activity.

The differential roles of the storage and comparator units can be analyzed further by examining the effect of removing them selectively. Removing a comparator unit from a minimal network has very little effect on the spatial pattern of the basins of attraction (Fig. 7*C*). Nonetheless, the matching process is disturbed, and there are several matching errors (data not shown). Removing a storage unit (Fig. 7*D*), on the other hand, substantially disrupts the spatial structure of the attractor basins. The values of the attractors are changed, their numbers are reduced, and several targets end up in the same basins of attraction.

The spatial patterns of attractor basins are similar in small and large networks (Fig. 7*A,B*). However, there are some differences. First, larger networks degrade more gracefully with lesioned storage units. Second, the boundaries between basins of attraction are more sensitive to starting state in the larger networks.

Comparison of model and empirical data

In the analysis of networks of relatively large size, two tests were required to show the function of the different types of hidden units: first, examination of the activation properties during the time course of a trial as a function of the stored and current input stimulus; and, second, the effect on the basins of attraction when the unit was deleted from the network. Because we are unable to perform the second of these tests on the experimental data, we are limited to the statement that a neuron is consistent with some particular functional type.

For the PM population, 68 neurons were included in the final data set. Only cells with unimodal directional tuning and increases in activity were included. Twenty-four of those 68 neurons had activity patterns that resembled, by qualitative assessment, the pure functional types found in the minimal model. Of these, there were seven storage units and 17 comparator units. The remainder showed either complex activity patterns that appeared to include elements of both functions or, in a few cases, were of types that did not appear in the model. For the PF population, 37 neurons were included using the same criteria. Of those, two resembled storage units, and five resembled comparator units. The remainder were of mixed or different types.

Storage units

The storage units in our model network have several characteristics: they are set to stimulus-specific levels of activity by the first stimulus in a trial, sustain this activity with little disruption despite the presentation of distractors, and are directionally

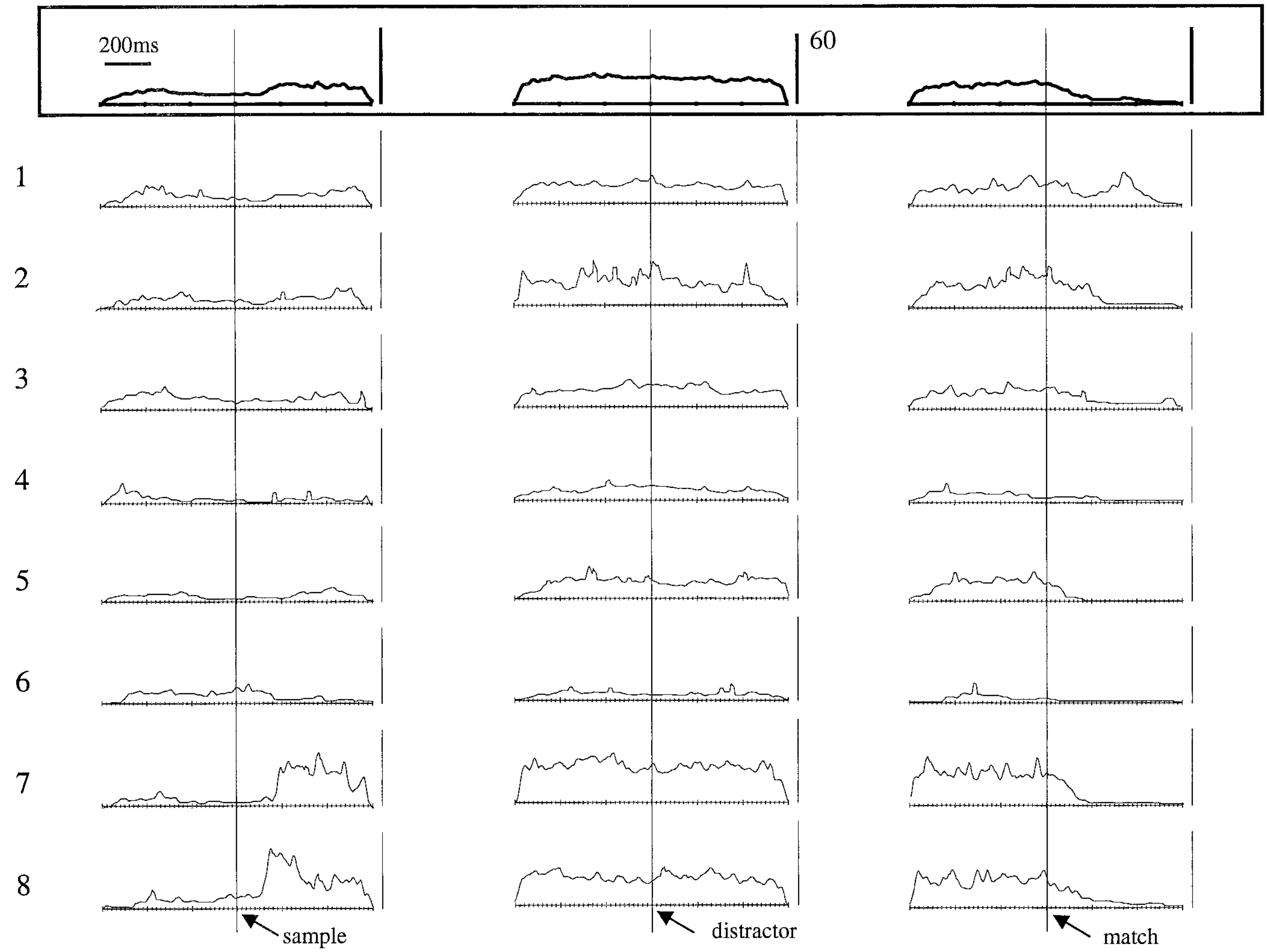

Figure 9. A storage unit from PM. Average neuronal activity for stimuli at each of the eight locations in the training set (*rows 1–8*). *Left column*, Activity aligned on the onset of the sample stimulus. *Middle column*, Activity aligned on the onset of a distractor stimulus. *Right column*, activity aligned on the onset of a match stimulus. Each *trace* is an average across a variable number of trials, plotted to the same scale (activity scale is in impulses per second). *Top row*, Average for each column.

tuned to the location of the initial, gated stimulus. Figure 9 shows the activity of a typical neuron with sustained activity. Illustrated are the periods before and after samples, distractors, and matches. The PM neuron shown in Figure 9 has the three characteristic response features of the storage units of the model. It is set to stimulus-specific levels of activity by the first stimulus in a trial (Fig. 9, *left column*); it sustains this activity with little disruption through the presentation of distractors (Fig. 9, *middle column*); and it is directionally tuned to the location of the initial, gated stimulus (Fig. 10). Unlike the storage units in the model, the activity of the cell diminishes soon after the match stimulus presentation (Fig. 9, *right column*). This characteristic is typical of neurons with delay period activity in a wide range of experiments (Zipser, 1991), and it can be reproduced in the model by gating in a stimulus representing the center location at the end of a trial (Fig. 4). The shutoff indicates that, in addition to some general source of input gating that loads new information into the system, there is also an information source, not addressed by our model, that indicates when the task is complete and information need no longer be maintained. The storage-like units in PF had similar characteristics.

Comparator units

Comparator units in the model have two critical properties that can be compared with PM and PF neurons. They have a directionally tuned response to stimuli presented during any stage of a trial, and their responses are modulated by the value currently stored in memory. In the model, the first of these properties results from comparator units receiving afferent information about the current input. The second property stems from storage unit input to comparator units. In the cortex, both PM and PF contain neurons that exhibit these two properties. The responses of a PM neuron to each of the eight input stimuli, whether that stimulus is a sample (Fig. 11, *left column*), a distractor (Fig. 11, *middle column*), or a match (Fig. 11, *right column*), gives a rough idea of the directional tuning of the cell. As with the comparator units of the model (Fig. 5), this neuron shows directional tuning to current stimuli, as well as responses that are modulated by stored values (Fig. 12). In PF cells, the directionally tuned response to the sample stimulus presentation tends to be similar to that of a current stimulus presentation. In contrast, the directional preferences of PM cells are more complex (see di Pellegrino and

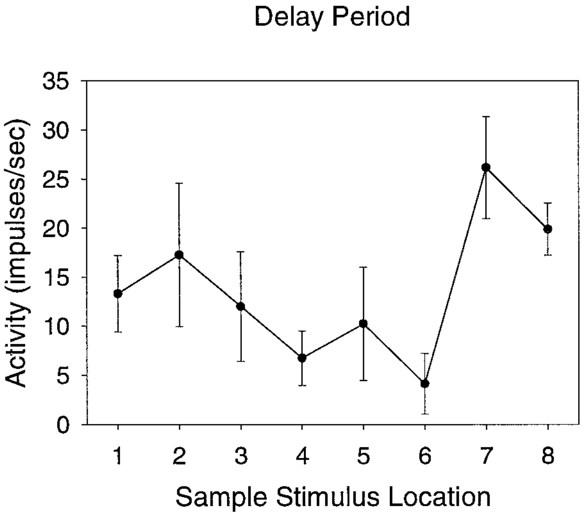

Figure 10. Tuning curve for PM cell illustrated in Figure 9. Activity and SE/SD for activity during the delay period (for the 750 msec period immediately preceding match stimulus onset) across stimulus location.

Wise, 1993b). In model neurons, the sample tuning curve is often quite similar to the current input tuning curve; however, there are many instances in larger networks in which the two curves are dissimilar.

Match output units

There were no match units in the recurrent layer of the model. The two principal characteristics of the match output unit, as expressed in the present terminology, are that it is "untuned" for direction; i.e., it reports a match regardless of the spatial location of the match stimulus, and it does not show activity at any other period during a "trial." By those criteria, there were no match units found in either PM or PF. Taken literally, this finding suggests that the actual decision that a match has occurred takes place in a different brain area, although all the relevant information is present in PM. However, closer examination of the data revealed a small number of units, four in PF and one in PM, with properties that resembled the match output neuron in the model more closely than did any of the hidden units of the model.

This characteristic of frontal cortex activity was explored further by examining the correlation between the depth and selectivity of directional tuning. We calculated s_i, d_i, a_i, and m_i for each model neuron (Eqs. 2–5) in both larger and minimal networks, as well as for PM and PF neurons, selected as described above. Figure 13 illustrates large and small values for the directional selectivity index and the depth-of-tuning index, s_i and d_i, respectively. s_i measures the extent to which activity in all nonpreferred directions deviates the maximal activity, and d_i measures the greatest proportional reduction from maximal activity. We found that PM, PF, and model data all show a significant, positive correlation between the depth (d_i) and selectivity (s_i) of directional tuning that is similar in certain details (Fig. 14). For activity after match presentation in larger networks, $r^2 = 0.877$ ($n = 50$); in PM neurons, $r^2 = 0.744$ ($n = 68$); and in PF neurons, $r^2 = 0.753$ ($n = 37$). Neither the independent correlations of PM and the model nor those of PF and the model significantly differ (two-tailed test, $\alpha = 0.01$) (Myers and Well, 1991).

There is an apparent difference between the model and neuronal distributions as the origin is approached (Fig. 14). More model units have very low depth-of-tuning as well as very low directional selectivity indices. For all of those model units, it was found that the level of modulation, m_i, was also very low, as was their activity during the delay period and after the match event. The very low activity levels (<6% of maximal activity) of these hidden units and the small weights linking these units to the network output suggest that these units contribute little, if any, to the operation of the model. Frontal cortex neurons with such low levels of activity and modulation are unlikely to be sampled with current neurophysiological sampling methods. Four PF cells and 1 PM cell with low d_i and s_i values were sampled (Fig. 14). However, unlike the hidden units of the model, these frontal cortex neurons have high modulation (m_i) levels (>44% of maximal modulation), with a moderate, rather than low, activity during the delay period. One of the PF neurons showed the maximal activity, which was a postmatch modulation of >200 impulses/sec, one of the highest levels of activity ever reported for frontal cortex. Thus, these four PF and one PM neurons resemble the match output cell of the model more closely than any of the hidden units of the model in having low directional tuning but high activity modulation after the match event. Some of these neurons also differ from the match output neuron in having significant activity after the sample stimulus, as well as the match stimulus.

DISCUSSION

Contribution of model

The model described here augments previous DMS models in several important ways. First, it incorporates the matching function in the same network as short-term information storage. Although active storage has been modeled (Lukashin, 1990; Zipser, 1991; Zhang, 1996), it has not previously been combined with matching in a single model. Second, the present model also allows for distractors and can distinguish repeats from true matches. Third, the model can be generalized readily to other short-term memory tasks, such as nonspatial DMS, nonmatching, and paired associate tasks. For example, the dimension and format of the input can be altered to represent the desired stimulus as x_{height}, x_{width}, *and* x_{color} for object x in a nonspatial DMS task.

We found that in models with a small number of hidden units, these units, to a first approximation, could be characterized as storage or comparator units. Recall, however, that the minimal network described above contains two storage units and six comparator units (Fig. 6). This suggests that comparator units are involved in storage, because two storage units cannot collectively represent eight different sample locations, and moreover, because some storage attractor basins are still intact after a storage unit lesion. Thus, although minimal-sized models produce predominantly storage or comparator units, these two unit types interact to support each other functionally. By contrast, networks with relatively large numbers of units tended to have what appear to be multifunction units. In this regard, the behavior of hidden units in large networks more closely resembles that of frontal cortex neurons. The role played by these multifunction units in the model is difficult to determine, as has proven to be the case for frontal cortex neurons. In the present analysis, we have examined the properties of frontal cortex neurons mainly in comparison with the properties of large networks. We focused on the properties of storage units, comparator units, and the match output unit, which signals the decision of the network.

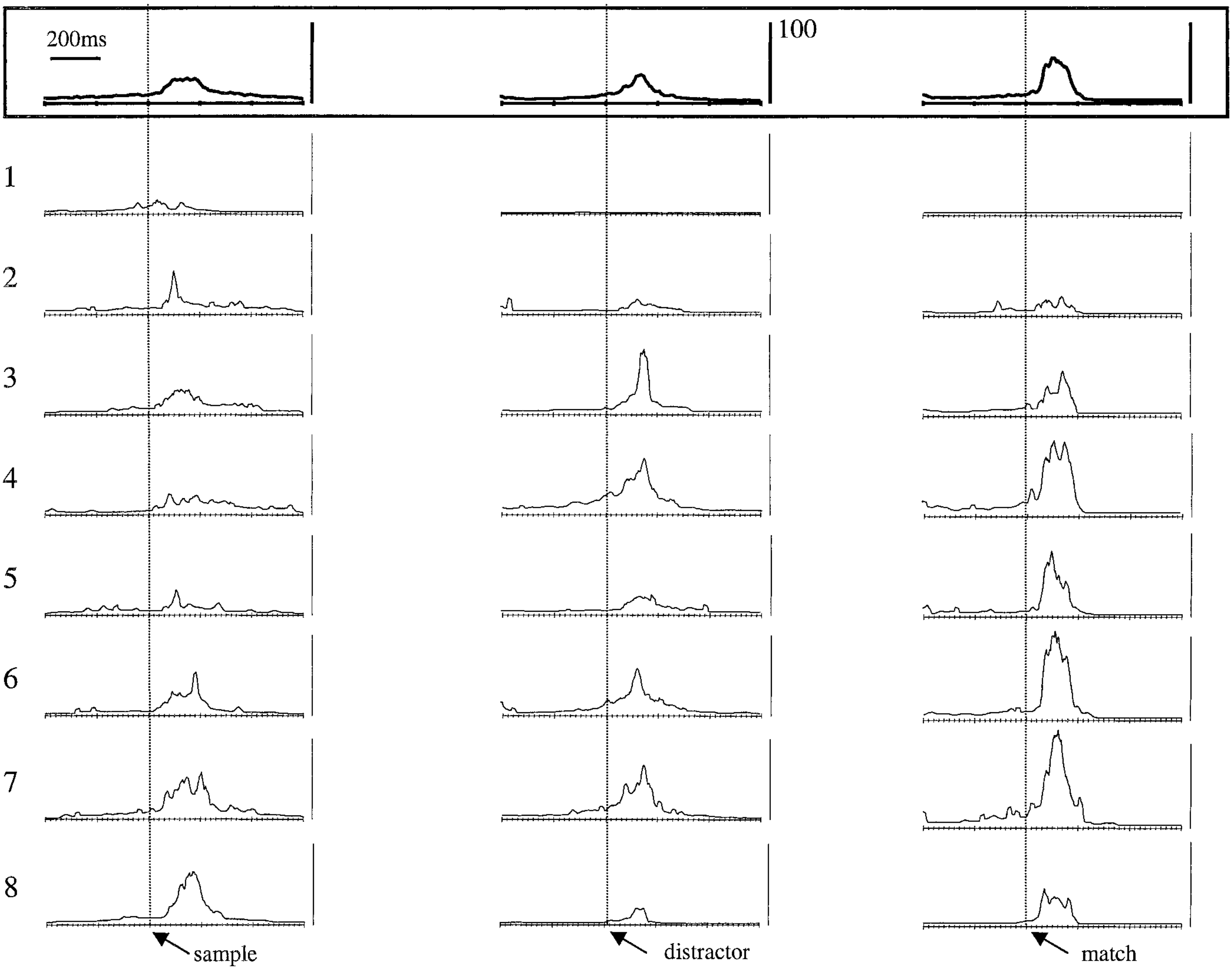

Figure 11. Comparator unit from PM. Format is as in Figure 9.

Storage units

Short-term active memory of the kind used in DMS must be buffered against distractor inputs. In the lateral intraparietal area and in areas 7a, IT, and V4, delay activity appears to carry information only about the immediately preceding stimulus (Mountcastle et al., 1987; Miller et al., 1993; Steinmetz and Constantinidis, 1995). In those sensory cortical areas activity is usually "reset" by distractor stimuli (Miller et al., 1996), presumably reflecting the displacement of sample information. In contrast, PF and PM neurons exhibit much less, if any, resetting by intervening distractors (di Pellegrino and Wise, 1993a; Miller et al., 1996). The buffering of stored information across distractors reinforces the conjecture that the loading of new information into active short-term memory requires some kind of specific gate or load signal. This signal appears to be necessary for shifting the units to a new attractor. For units in which memory is reset to a new value for each new stimulus, such a load signal would not be required. For those units, the afferent stimulus could shift the network to a new attractor. The source of the load signals for short-term memory is not known. One possibility is what has been termed anticipatory or precue activity: a directionally nonselective signal that precedes a temporally predictable event (Mauritz and Wise, 1986; Vaadia et al., 1988). In most tasks, the sample in a DMS task is predictable in this sense. In principle, this signal can provide information about the control of the sequence of computations that generate continuous cognition. The same reasoning also applies to the signals that turn off sustained activity at the end of a trial.

In the model, the function of the storage units in maintaining information about the location of the sample is unambiguous. However, in animal behavior, the situation is not always as clear. During periods when animals must remember a stimulus location, they may also reorient selective spatial attention to the same location and/or, under certain circumstances, maintain an intention to make a movement, either oculomotor or skeletomotor, to that location. It is also possible that animals could use overt movements for self-cueing, for example, if an animal looks in the direction of the sample stimulus or adopts a posture that could encode the same information. Eye position and electromyograpic measurements seem to exclude overt motor strategies for the matching task (di Pellegrino and Wise, 1993a), but covert actions cannot be ruled out. For the present purpose, these distinctions are not particularly pertinent. Whatever the strategy of the animal for retaining information about the sample stimulus, the require-

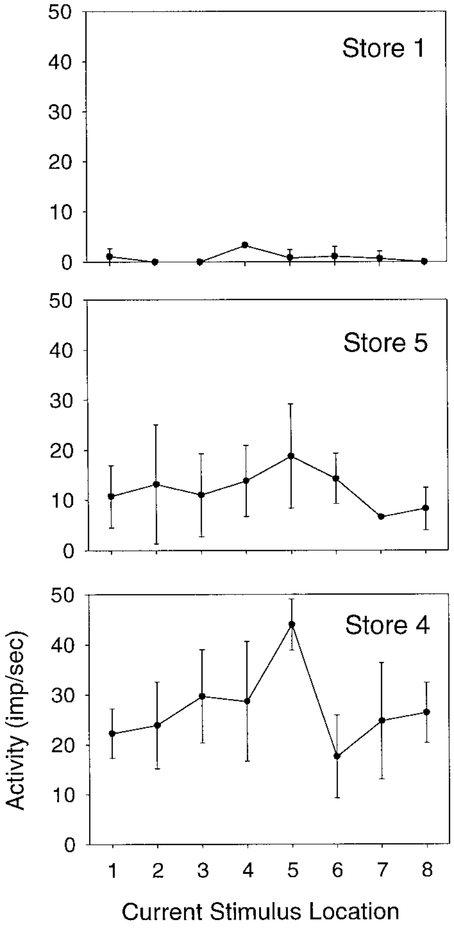

Figure 12. Tuning curves for PM cell illustrated in Figure 11. As in Figure 5 from the model, each plot shows the response to stimuli at one of the eight locations in the training set after a sample at one location. *Top, middle, bottom*, Sample stimulus at locations 1, 5, and 4, respectively.

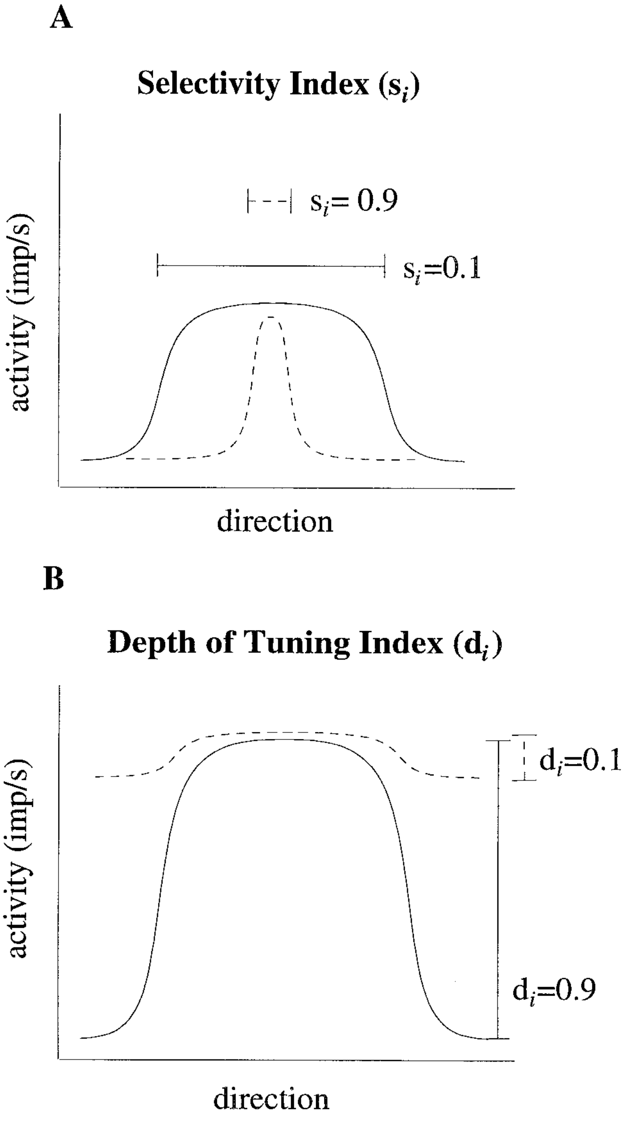

Figure 13. Schematic of tuning properties described by the directional selectivity (s_i) and depth-of-tuning (d_i) indices (see Eqs. 2, 3).

ments of the DMS network remain much the same: location-specific sustained neural activity and comparison between this stored information and current stimuli.

We did not examine the dynamics of cortical discharge rates in detail. It has been noted that neuronal activity in a delay period has many patterns, including, in frontal cortex, a buildup of delay period activity over time (Quintana and Fuster, 1993; Miller et al., 1996). Although the model, in its current form, cannot account for such augmentation in discharge rate, a minor modification that enhances the effect of the previous network activation is likely to produce such an effect.

Comparator units

The model network functions by virtue of the combinations of inputs from current stimuli and storage units onto comparator units. This conclusion is supported by three lines of evidence: the properties of comparator unit activity, the synaptic weights that these neurons have on the match output neuron, and the effects of selective removal of comparator neurons from a trained model. Comparator unit activity reflects both the location of the sample stimulus (as provided by the storage units) and the location of the current stimulus. It appears that a unique combination of storage and current stimulus inputs to each comparator unit is used by the network to identify match events. This conclusion is supported further by the fact that the largest synaptic weights on the match output neuron of the model arise from the comparator cells as well as from the observation that removal of comparators from the network causes many "reporting" errors without changing the basins of attraction.

Of course, it is impossible to perform the latter two analyses on the frontal cortex neurons. No one knows the weight of its neurons on output elements, and there is no method for ablating comparator units selectively in the frontal cortex. However, we can examine the activity patterns and levels of PF and PM neurons in the context of the behavior of hidden units in both minimal and relatively large networks. A large class of PF and PM neurons has the properties predicted by the model for comparator units. Comparator cells are characterized by receiving input regarding the stored stimulus, as well as the current stimulus. In terms of our neuronal activity indices (Eqs. 4, 5), this corresponds to moderate but nonzero delay period activity ($\bar{\imath}_{\text{delay}}$) and moderate postmatch stimulus modulation (m_{i}). PM, PF, and the model all contain populations of comparator cells by these criteria.

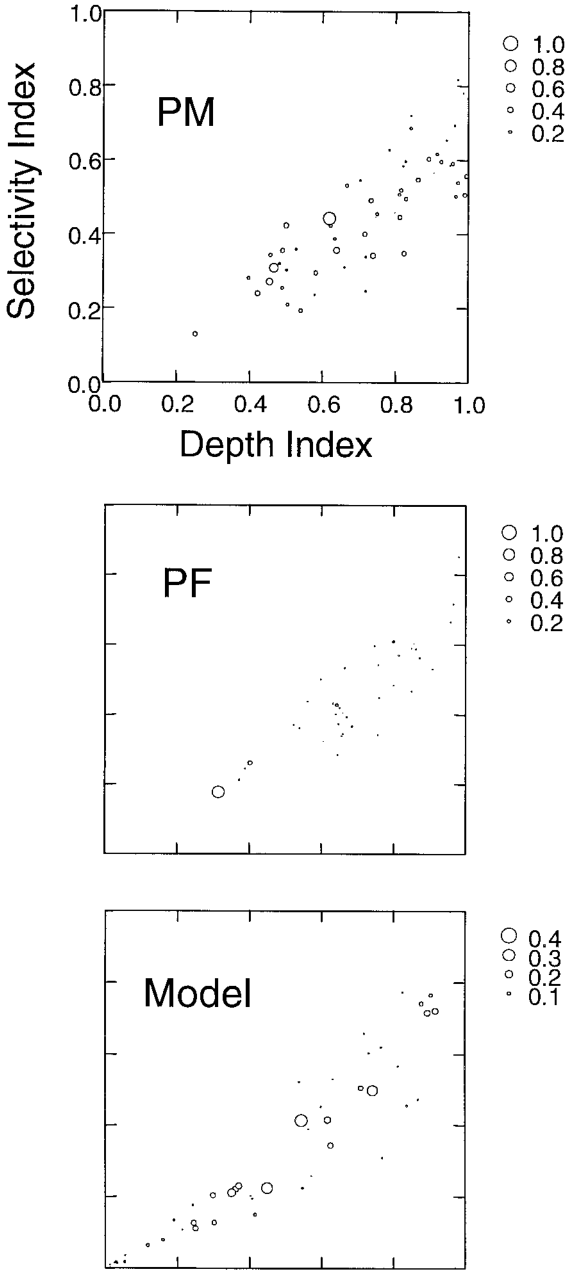

Figure 14. Directional selectivity index versus depth-of-tuning index for PM neurons, PF neurons, and model neurons. The diameter of each data *point* is proportional to the normalized activity modulation (m_i) after the match stimulus event. Because of the unusually large modulation of activity for one PF neuron (>200 impulses/sec), the rest of the PF population appears to have minimal modulation. However, this is merely a result of normalization to the maximum value.

Match output units

There were no true match units, i.e., units that became active only when a match occurred and did so equally for all match-stimulus locations, in PM, PF, or the hidden layer of the model. This negative finding is of importance, because it indicates that hidden units of the model probably do not make the match decision. In this context, we examined the behavior of the PF and PM neurons in our sample closely. The finding of four PF neurons and one PM neuron that had virtually untuned, but substantial, activity modulation after the match event suggests a larger role of frontal cortex neurons in the match decision than envisioned for the hidden units in the model.

Neural systems identification and top-down engineering

Top-down engineering concepts are useful in combined modeling and neurophysiological studies, especially as a source of conjectures concerning how a network may be solving a complex input–output mapping. As outlined in the introductory remarks, an optimal top-down solution to the DMS task could consist of an attractor network to store the stimuli and a match filter to detect matches. In the NSI model, the storage units implement the attractor network, and the comparator units implement the match filter. Thus, the NSI model did implement both of the proposed mechanisms, although it was not constrained to do so. Furthermore, the model network acquired the property of buffering of the relevant information, without individual elements exclusively devoted to this function, and lacked "pure" match units. These properties accord with the experimental results, although in an explicitly engineered system, components might have been designed with those properties. It is of interest that, among the many potential designs for a DMS network, the NSI model and perhaps the cortical network adopt the optimized architecture proposed in the top-down engineering solution described above.

REFERENCES

Amit DJ, Brunel N (1995) Learning internal representations in an attractor neural network with analogue neurons. Network 6:359–388.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE (1997) Temporal dynamics of brain activation during a working memory task. Nature 386:604–608.

Courtney SM, Ungerleider LG, Kiel K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608–611.

Cowan JD (1972) Stochastic of neuroelectric activity. In: Statistical mechanics (Ricce SA, Fread KF, Light JC, eds), pp 181–182. Chicago: Chicago UP.

Crammond DJ, Kalaska JF (1989) Neuronal activity in primate parietal cortex area 5 varies with intended movement direction during an instructed-delay period. Exp Brain Res 76:458–462.

di Pellegrino G, Wise SP (1993a) Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. J Neurosci 13:1227–1243.

di Pellegrino G, Wise SP (1993b) Effects of attention on visuomotor activity in the premotor and prefrontal cortex of a primate. Somatosens Mot Res 10:245–262.

Evarts EV, Tanji J (1976) Reflex and intended responses in the motor cortex pyramidal tract neurons of monkey. J Neurophysiol 39:1069–1080.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J Neurosci 2:1527–1537.

Georgopoulos AP, Kettner RE, Schwartz AB (1988) Primate motor cortex and free arm movements to visual targets in three dimensional space. II. Coding of the direction of movement by a neuronal population. J Neurosci 8:2928–2937.

Lukashin AV (1990) A learned neural network that simulates properties of the neuronal population vector. Biol Cybern 63:377–382.

Mauritz K-H, Wise SP (1986) The premotor cortex of rhesus monkeys: neuronal activity before predictable environmental events. Exp Brain Res 61:229–244.

Mikami A, Kubota K (1980) Inferotemporal neuron activities and color discrimination with delay. Brain Res 182:65–78.

Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 13:1460–1478.

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci 16:5154–5167.

Moody SL, Zipser D (1998) A model of reaching dynamics in primary motor cortex. J Cognit Neurosci 10:35–45.

Mountcastle VB, Motter BC, Steinmetz MA, Sestokas AK (1987) Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. J Neurosci 7:2239–2255.

Myers JL, Well AD (1991) Research design and statistical analysis, p 480. New York: Harper Collins.

Oppenheim AV, Willsky AS, Nawab SH (1996) Signals and systems, pp 170–172. Upper Saddle River, NJ: Prentice-Hall.

Quintana J, Fuster JM (1993) Spatial and temporal factors in the role of prefrontal and parietal cortex in visuomotor integration. Cereb Cortex 3:122–32.

Sanger TD (1994) Theoretical considerations for the analysis of population coding in motor cortex. Neural Comput 6:29–37.

Steinmetz MA, Constantinidis C (1995) Neurophysiological evidence for a role in posterior parietal cortex in redirecting visual attention. Cereb Cortex 5:448–456.

Steinmetz MA, Connor CE, Constantinidis C, McLaughlin JR (1994) Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. J Neurophysiol 72:1020–1023.

Tanji J, Taniguchi K, Saga T (1980) Supplementary motor area: Neuronal response to motor instructions. J Neurophysiol 43:60–68.

Vaadia E, Gottlieb Y, Abeles M (1982) Single-unit activity related to sensorimotor association in auditory cortex of a monkey. J Neurophysiol 48:1201–1213.

Vaadia E, Kurata K, Wise SP (1988) Neuronal activity preceding directional and nondirectional cues in the premotor cortex of rhesus monkeys. Somatosens Mot Res 6:206–230.

Weinrich M, Wise SP (1982) The premotor cortex of the monkey. J Neurosci 2:1329–45.

Williams RJ, Zipser D (1995) Back-propagation: theory, architectures, and applications. In: Developments in connectionist theory (Chauvin Y, Rumelhart DE, eds), pp 433–486. Hillsdale, NJ: Erlbaum.

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J Neurosci 16:2112–2126.

Zhou Y-D, Fuster J (1996) Mnemonic neuronal activity in somatosensory cortex. Proc Natl Acad Sci USA 93:10533–10537.

Zipser D (1991) Neural mechanism of short-term active memory. Neural Comput 3:179–193.

Zipser D (1992) Identification models in the nervous system. Neuroscience 47:853–862.

Zipser D, Andersen RA (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature 331:679–684.