Chapter 11

Learning and Recalling Objects with Naturally Varying Shapes

Computer simulations described in the previous chapter demonstrated that the putative visual-cognitive system can learn and recall canonically shaped objects that are presented on the normal foveal axis or at eccentric locations in the visual field. Furthermore, the tests provided evidence that recognition performance of the model is robust under conditions of visual noise and rotational transformation. In the simulations described in this chapter the proposed centroid-finding mechanism works with the synaptic matrix to enable the system to fixate and learn visual stimuli that are not canonical and presented at random locations in the visual field. Figure 11.1 is a block-flow diagram summarizing the processing sequence. The mechanisms within the labeled blocks embody the neuronal properties described in earlier chapters.

The performance of the model with noncanonical stimuli was tested by presenting it with handprinted lowercase letters, freely written on a digitizing tablet as one might write in a notebook but without the usual guiding lines. Handprinted lowercase letters are well suited as stimuli because each sample conserves the distinguishing underlying characteristics of the letter it represents and at the same time varies from one instance to another in its exact shape, size, angular orientation, and location with respect to a fixed reference axis (the normal foveal axis). The first test required the model network to learn and recognize letters printed by one person (Trehub 1990). In a second test, the system was given the task of learning and recognizing letters printed by five different individuals. Clearly, in the latter case, one would expect a much wider range of variation in shape, size, and slant over the sample of stimuli. Although performance of the network would be enhanced if each letter presented to the retina were normalized for variation in size and angular orientation, none of the simulations embodied size and rotation transformers to correct for such variations.

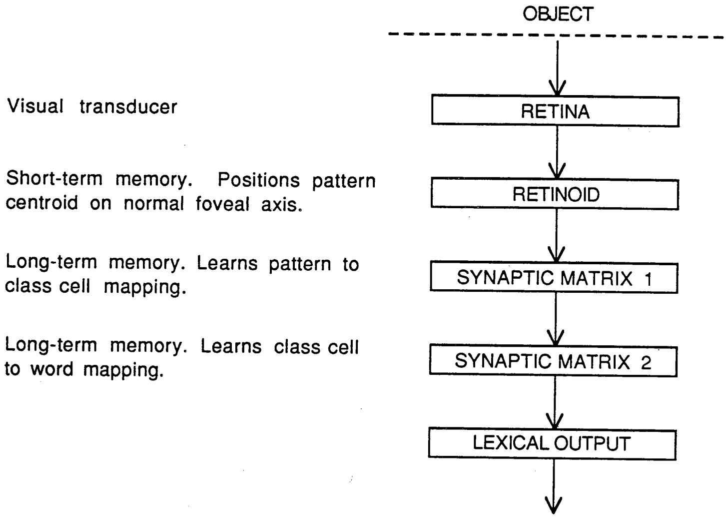

Figure 11.1
Block-flow diagram of simulated pattern recognition subsystem.

Test Procedure 1

A neuronal subsystem consisting of a retina, retinoid, synaptic matrix, and afferent aperture circuit was simulated in a digital computer. The gradient coefficient of dendrodendritic transfer to neighboring cells in the mosaic cell array was arbitrarily set at 0.5. The constant b (representing the initial transfer weight of filter cells) was disregarded because it is assumed to be very small and uniform for all such cells in the synaptic matrix. The constant c (representing the axon transfer factor) was set at 2.0. The coefficient k (representing the dendrite transfer factor) was set at 38.0. Error tolerance (ET) for centroid alignment on the normal foveal axis was set and held constant at 3 retinoid units.

Input to the retina was provided by a graphics digitizing tablet. Lowercase letters were handprinted in normal size on an unlined sheet of paper that overlay the digitizing tablet, which converted the pressure of the pencil trace to a spatially conforming digital signal pattern. This two-dimensional binary pattern constituted the stimulus to a 15×15–cell retina in the simulated network.

During the initial learning phase, each letter of the alphabet was presented one at a time. Because the letters were freely drawn within the fixed visual field of the system, most stimuli happened to project to parafoveal positions on its retina. Where the fixation tolerance was exceeded, the retinoid mechanism attempted to translate its representation of the letter pattern to the normal foveal axis before the letter was learned. Despite this effort to fixate each stimulus representation on the standard learning axis, the coarse resolution of the retina resulted in an inability to satisfy error tolerance for horizontal and/or vertical axes in about half of the stimulus presentations. When this happened, the model printed the message "CANNOT FIXATE THIS OBJECT," and another exemplar of the same letter was printed. This phase of the procedure was completed after each letter of the alphabet had been fixated once and then learned, which resulted in the synaptic "tuning" of 26 filter cells in the detection matrix. The discharge of any class cell was then taken to indicate the presence of the letter previously learned by its synaptically coupled filter cell.

In a correction phase following the initial learning sequence, letters were printed, one at a time, starting from the beginning of the alphabet. The model was required to fixate and recognize (by appropriate class cell discharge) the letter presented. If the response was correct, the next letter was written. If the response was wrong, the misidentified letter was learned (that is, an additional filter cell was synaptically modified) as another exemplar of its stimulus class. The model was then required to recognize another printing of the same letter. If the response was correct, the next letter of the alphabet was presented; if incorrect, the current letter was learned as a new exemplar of its class. This procedure was repeated until 33 lowercase exemplars had been learned, at which point recognition performance was tested without correction.

In the uncorrected test sequence, each letter of the alphabet was presented for fixation and recognition. Each stimulus-response pair was recorded for all 26 letters, and the entire procedure was repeated through five runs of the alphabet. Thus, 130 letters were drawn and 130 recognition responses were made with no corrective learning. Similar uncorrected test sequences were run after 36 and 46 exemplars had been learned in interspersed correction phases.

Upon completion of the learning routines, the distributions of synaptic transfer weights (ϕ) on 46 filter cell dendrites had changed in accordance with the basic learning mechanism. Figure 11.2 shows the synaptic profiles of a few selected filter cells in the detection matrix together with the alphabetic character associated with each learning-induced synaptic distribution. The selectivity of recognition response is determined by the differences among such ϕ distributions over the population of filter cells in the detection matrix.

The graph in figure 11.3 shows the improvement in recognition performance in the uncorrected test sequences as the number of

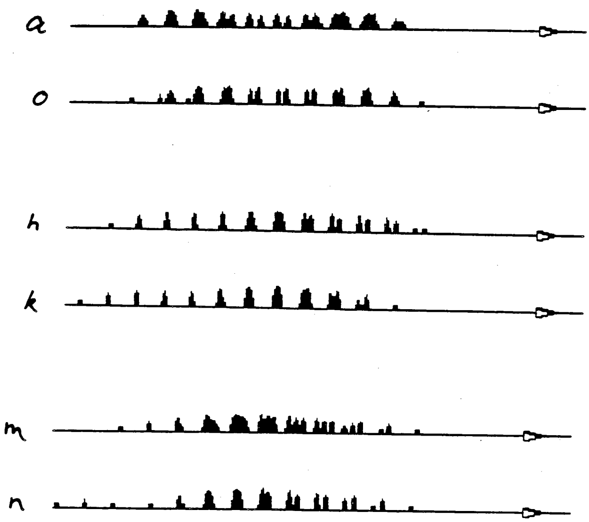

Figure 11.2
Examples of the distribution of synaptic transfer weights (ϕ) on the dendrites of filter cells that learned handprinted letters in test procedure 1. Each point on the dendritic line represents the relative magnitude of transfer weight for that synapse. Letters learned are shown to the left of each filter cell. Source: Trehub 1990. Copyright Lawrence Erlbaum Associates, Inc. Reproduced by permission.

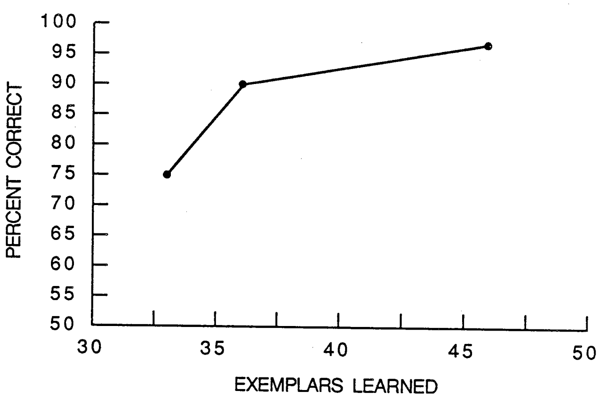

Figure 11.3
Graph showing the percentage of correct letter recognition responses during the test trials plotted against the total number of exemplars learned before each test. Source: Ibid. Copyright Lawrence Erlbaum Associates, Inc. Reproduced by permission.

learned exemplars increased from 33 to 46. Each point on the graph is based on 130 stimulus-response pairs. When 33 exemplars were learned, response was 75 percent correct. With 36 learned exemplars, response was 90 percent correct. After 46 exemplars had been learned, response was 97 percent correct.

Confusion matrices based on all stimulus-response pairs in the initial uncorrected test sequence and in the last test sequence are presented in figures 11.4 and 11.5. For most letters, confusions occur when the stimulus and response share similar graphic features.

Test Procedure 2

In the second test, the task of learning to identify handprinted letters was made more difficult: the network was required to learn and recognize lowercase letters printed by five different individuals, and it was not permitted to reject stimuli with centroids that it could not align within a tolerance of plus or minus three retinoid units from the normal foveal axis. Thus, unlike the previous simulation, an attempt at recognition was required for each stimulus presentation.

The neuronal subsystem consisted of a retina, a retinoid, and a synaptic matrix, with the retina dimensioned at 16 $\times$ 16 cells. The gradient coefficient for dendrodendritic transfer to neighboring cells in the mosaic cell array was set at 0.6. The axon transfer factor (c) was set at 1.0 and the dendrite transfer factor (k) at 100.

Test stimuli were obtained from the writing of five individuals. In the first sample, each writer printed all lowercase letters of the alphabet in three complete sequences on the page of a notebook, providing an initial total of 78 stimuli per individual (figure 11.6). In the second sample approximately two months later, each writer again provided three complete printings of the alphabet.

Stimulation to the retina was provided by digitized tracings of the handprinted letters. Most letters projected to parafoveal positions on the retina as before, but in this case, instead of rejecting stimuli that could not be fixated within a constant error tolerance, the system worked to minimize fixation error before attempting to recognize each stimulus. It first tried to align the stimulus centroid to within plus or minus two retinoid units of the normal foveal axis. If quadrantal disparities exceeded this limit, the model relaxed its error tolerance by one unit (to plus or minus three units). This process was iterated until the minimum tolerance was established within which fixation could occur (figure 7.4). At this point, the letter pattern was gated through to the synaptic matrix to be recognized, and if the response happened to be incorrect, that stimulus was learned.

Response to Handprinted Alphabetic Characters

Recognized as

Letter Presented	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	3																									2
b		5																								
c			4												1											
d				4														1								
e					5																					
f						3			1			1														
g							5																			
h		3						2																		
i									5																	
j				3						2																
k								1			3			1												
l						2						3														
m													4										1			
n													2	3												
o															3						2					
p																3		2								
q																	5									
r					1													4								
s															2				3							
t												1								4						
u				1																	4					
v																					1	4				
w								1															4			
x																								5		
y				1											1										3	
z																										5

Number of Exemplars = 33 Correct Responses = 75%

Figure 11.4
Confusion matrix following the learning of 33 exemplars. Source: Ibid. Copyright Lawrence Erlbaum Associates, Inc. Reproduced by permission.

Response to Handprinted Alphabetic Characters

Recognized as

Letter Presented

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	5																									
b		5																								
c	1		4																							
d				5																						
e					5																					
f						4										1										
g							5																			
h								5																		
i									5																	
j										5																
k											5															
l												5														
m													5													
n														5												
o															5											
p																5										
q							1										4									
r																		5								
s																			5							
t																				5						
u																					5					
v																					1	4				
w																							5			
x																								5		
y																									5	
z																										5

Number of Exemplars = 46 Correct Responses = 97%

Figure 11.5
Confusion matrix following the learning of 46 exemplars. Source: Ibid. Copyright Lawrence Erlbaum Associates, Inc. Reproduced by permission.

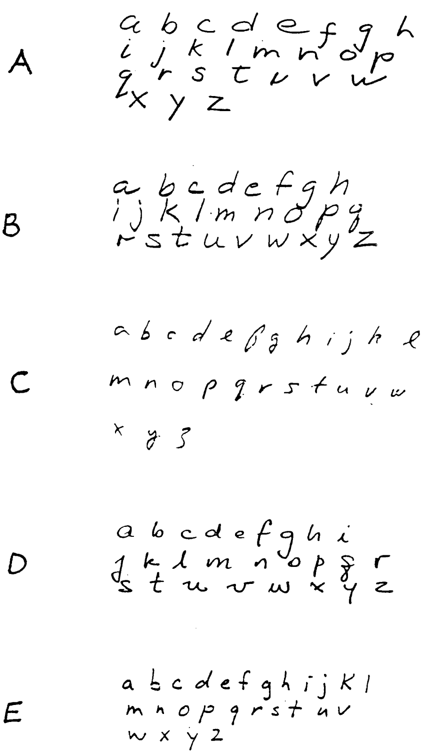

Figure 11.6
Samples of the handprinted stimuli produced by each of the five writers in test procedure 2.

All of the letters from the first sample provided by individual A were presented to the model. At each presentation, if the response was correct, the next letter of the alphabet was presented; if the response was wrong, the misidentified letter was learned, and then the next letter was presented. This procedure was repeated until all letters provided by source A in the first sample had been presented. Stimulus presentation, recognition response, and learning (if an error was committed) continued in the same fashion for the letters from each source until all stimuli in the first sample had been shown. Thus, three successive runs of the lowercase alphabet printed by each of five individuals were presented to the network. At this point, the entire procedure was repeated with the second sample of letters. In all, 780 stimulus-response trials were obtained in this simulation.

Figure 11.7 shows the synaptic transfer weight (ϕ) profiles of the

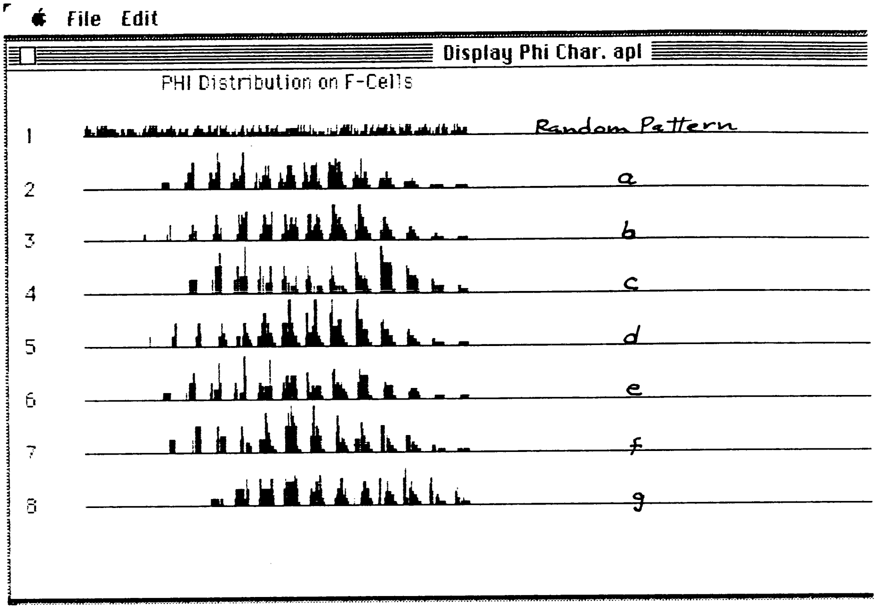

Figure 11.7
Synaptic transfer weights on the dendrites of the first eight filter cells modified during learning in test procedure 2. The synaptic profile on filter cell 1 is the result of learning a random visual pattern.

first eight filter cells that were modified during the learning procedure and the alphabetic character associated with each.

The proportion of correct responses made by the model for each complete presentation of the lowercase alphabet (the number of correct identifications divided by 26) was plotted against the number of exemplars that had already been misidentified and learned just before each new test of the alphabet began. These data are shown in figure 11.8 as percentage correct against exemplars learned. There was a generally progressive improvement in the model's performance as it was exposed to more stimuli despite the fact that for each character there were natural variations in size and orientation as well as shape, the model was not permitted to reject stimuli with centroids that could not be aligned within narrow error tolerance, and the letters to be identified were produced by five individuals with clearly different writing styles.

For the inputs in the second sample, when the model had to identify the writing of a given individual, it had already learned something about the writing of four other individuals. One might wonder

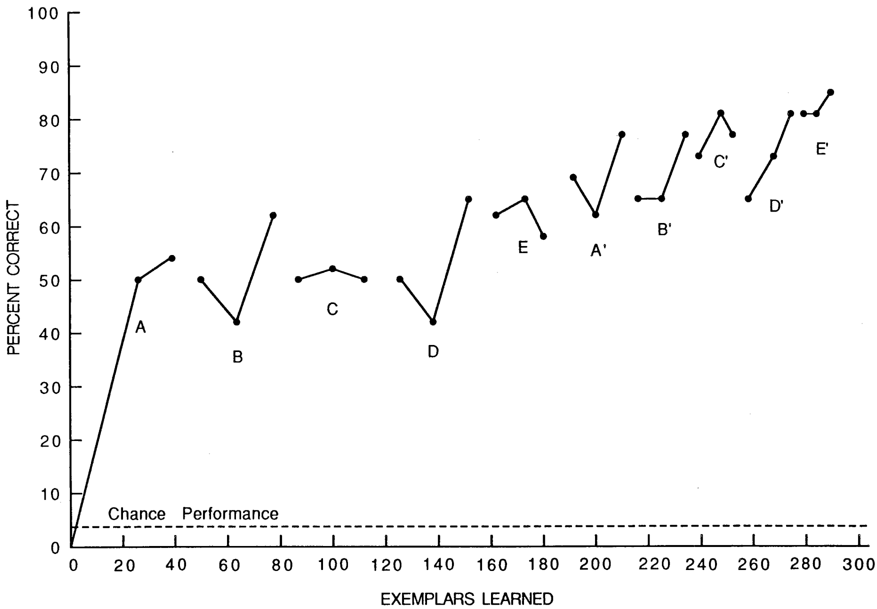

Figure 11.8
Graph showing the percentage of correct letter recognition responses over each run of the alphabet presented by each of the five writers, plotted against the number of exemplars learned at the start of each test run. Letters A–E indicate stimuli produced by the different writers. Letters without primes indicate sample 1. Primed letters indicate sample 2.

if the addition of filter cells tuned to the differing letter characteristics of other writers would interfere with the recognition of new exemplars. If, indeed, this is the case, then table 11.1 shows that any interference effects that might exist are more than offset by a positive generalization effect from the exposure to other exemplars, which results in a substantial improvement in performance. The gain in the percentage of correct identifications from the first to the second sample over the five stimulus sources ranges from 18 to 34 percent, with an overall average gain of 24 percent.

Comments

In these tests, the model does not depend on feature extraction or feature processing for learning or recognition. This differs significantly from models requiring preprocessed critical features for their input or innate "feature detectors" (Hinton 1981). The system simply

Table 11.1
Change in correct recognition responses

	Percentage Correct		
Writer	Sample 1	Sample 2	Percentage Gain
A	35	69	34
B	51	69	18
C	51	77	26
D	52	73	21
E	62	82	20

Average gain = 24%

learns its current pattern of retinal excitation after it has been shifted on the retinoid so that the pattern centroid is as close as possible to the normal foveal axis, given the physical constraints of the system. Objects that subtend large visual angles will naturally be partitioned by the visual apparatus before they are learned one part at a time, whereas small objects like alphabetic characters may be learned at once as holistic patterns.

Given the coarse resolution of the retina in the simulation, the lack of compensation for variations in size and slant, and substantial variability in the shape of each handprinted letter, the level of recognition performance exhibited by the model suggests that natural pattern recognition can proceed very well without prior feature extraction when the objects to be recognized are small.

These results also bear on the issue of representation by an average-based prototype versus representation by a subset of exemplars. The system modeled here does not compute a single prototypical distribution of synaptic transfer weights to represent a given letter of the alphabet. Instead, filter cells are separately tuned to different exemplars of each object (letter) class. The robust generalization exhibited in the simulation tests follows naturally from the intrinsic ordinal logic of the synaptic matrix. This mechanism ensures that even in the absence of a good match between an input pattern and its appropriate filter cell, if there is no better match with an inappropriate cell, the correct recognition response will occur.

The mode of internal representation used by the putative brain system (storage of exemplar instances in the form of distinct synaptic weight distributions on distinct filter cells) can explain another aspect of letter recognition: we are often able to identify the person who did the writing. This would not be possible if all instances of a particular object were represented by some normative prototype. When exemplars are stored, however, one can associate the name of the person producing a letter with the individual characteristics of the pattern that is learned. In a more general and important sense, when objects are represented discretely in the brain, any contextual events that might be significant can be associated with a particular experiential instance.

A fundamental question that can be asked of any quantitative model concerns the range of variation its parameters can tolerate before the model loses its effectiveness. The narrower the tolerable range is, the less robust is the model. In the simulations described here, two different sets of values were assumed in the learning formula. For the first simulation, the coefficient for dendrodendritic transfer was 0.50, the axon transfer factor (c) was 2.0, and the dendrite transfer factor (k) was set at 38.0. In the second simulation, the corresponding values were 0.60, 1.0, and 100. The fact that recognition performance was satisfactory with either set of parameter values is additional evidence of the robustness of the postulated learning mechanism.