Inter-trial neuronal activity in inferotemporal cortex: A putative vehicle to generate long term visual associations.

Volodya Yakovlev[1], Stefano Fusi[2,5], Elisha Berman[3], and Ehud Zohary[1,4]
[1] Dept. of Neurobiology, Institute of Life Science, [2] Racah Institute of Physics, [3] Human Biology Research Center, Hadassah University Hospital, [4] Center for Neural Computation, Hebrew University, Jerusalem, 91904, Israel. [5] INFN, Università di Roma "La Sapienza", Italy

Corresponding Author:

Dr. Ehud Zohary

Dept. of Neurobiology, Institute of Life Science,
Hebrew University, Jerusalem, 91904, Israel.
Tel: 972-2-6586-737
Fax: 972-2-6586-296
Email: udiz@lobster.huji.ac.il

Abstract

A population of neurons in anterior inferotemporal cortex has sustained activity following the presentation of specific visual stimuli when monkeys perform a delayed match-to-sample (DMS) task. Typically, only stimuli that are repeatedly shown elicit robust delay activity. When the sample stimuli were shown in a fixed temporal order, the few images that evoked delay activity in the same neuron were often neighboring stimuli in the sequence. Therefore, this delay activity was suggested as the neural correlate of associative long-term memory of visual images. We report here that stimulus selective sustained activity is evident also following the presentation of the test stimulus in the same DMS task. We demonstrate, using a neural network model, that the persistence of stimulus selective activity across the inter-trial-interval can lead to similar menmonic representations (distribution of delay activity across the neural population) for neighboring visual stimuli. Thus, the neural machinery for the generation of long-term stimulus-stimulus associations may exist in inferotemporal cortex.

Introduction

Most of us can remember the next melody on a record once the current tune is over, or recall the alphabet in its correct sequential order. These are classic examples of generating an association between stimuli which have been presented in a fixed temporal order. A visual example of such a phenomenon may be the way we navigate in an environment we are not well acquainted with. In such circumstances we typically use remembered snapshots of visual scenes to verify that we are on the correct track, with one serving as a cue leading us to the next expected landmark.

Our current understanding of the neuronal basis for the formation of such a long-term associative memories is only marginal. An important step towards deciphering the neuronal basis of long-term associative memory was carried out by Miyashita and his colleagues. Miyashita [1] trained monkeys to perform a delayed match-to-sample (DMS) task which required the monkey to remember the identity of a sample stimulus during a (16 sec.) delay interval, and respond differently if the following test stimulus was identical (in the "match" condition) or different (in the "non-match" condition) from the sample stimulus. The novelty in his experimental design was that the sample stimuli were presented in a fixed temporal order. After the monkeys were highly experienced with the task, he recorded the activity of single neurons in inferior temporal (IT) cortex, during the delay period between the presentation of the sample and test stimuli. Miyashita found that some IT neurons had enhanced levels of firing rates throughout the delay interval long after a specific sample stimulus was presented, as has been reported by others recording in IT and prefrontal cortex [2, 3, 4, 5, 6]. Although the monkey could preform the DMS task using novel stimuli as well, only stimuli that were highly familiar evoked this delay activity. An important finding was that the few visual stimuli that generated delay activity in the same IT neuron, were more likely to have been nearest neighbors in the fixed temporal sequence during the training period. This was in spite of the fact that the temporal order of the sample stimuli was totally irrelevant for performing the DMS task. This aspect of the neuronal response led Miyashita to suggest that "the selectivity acquired by these cells represents a neuronal correlate of associative long-term memory of pictures".

Recently, Amit and colleagues proposed a comprehensive theoretical framework for understanding the development of associative long term memory, based on the phenomena observed by Miyashita [7, 8]. According to this approach, the sustained delay activity is a feature of the pattern of connectivity between neurons, rather than a feature of a single neuron: The persistent delay activity is maintained by recurrent synaptic feedback between interconnected neurons within a local module, built up as stimuli become familiar. The memory process is initiated by presentation of the visual stimulus, which generates a pattern of response across the neuronal population. Following removal of the visual stimulus, due to the feedback connections within the neuronal population, the dynamics of the network is such that it settles into a stable state (the attractor), in which most neurons are firing at their spontaneous level, but some distinct neurons continue firing at elevated levels although the visual stimulus is no longer present. The stable state implies that this pattern of firing continues until a new afferent input (from a new, effective visual stimulus) changes the state of the network components. Since each visual stimulus evokes a characteristic pattern of delay activity, the delay activity distribution is the neuronal engram of the last familiar stimulus seen. The distributed nature of the representation allows storage of a large number of patterns (i.e stable delay activity distributions) in the same neural module, by the same synaptic structure.

One key property of such a network is its pattern completion abilities: The distributed representation across a large neuronal population, makes it relatively immune to noise. If the pattern of activity during the presentation of a modified or degraded visual stimulus has some resemblence to the pattern evoked by the original stimulus, the network will reach the same neuronal delay activity pattern (i.e. the dynamics will flow toward the same attractor). This type of a neuronal behavior, i.e. IT delay activity which is immune to moderate levels of noise in a visual stimulus, has been reported in [9].

It is important to note that the stable attractors are formed during a slow learning process which shapes the synaptic structure between the network members. Therefore, delay activity should be evident only for stimuli which have been repeatedly presented to the animal, as has been found by Miyashita [1]. The memories are embedded in the synaptic structure through an unsupervised Hebbian learning rule. Thus, no special assumptions or requirements are needed to generate the required synaptic structure.

Last, and most important, this framework can lead to an association betweem stimuli repeatedly presented in temporal proximity because the delay activity can link events separated in time: Neurons that are part of an attractor of one stimulus, will remain active during the delay period, until the presentation of the next stimulus. This joint activity (within a time window of tens of msec) will allow for Hebbian strengthing of the synapses between neurons belonging to the two populations. If the stimuli are systematically presented in a fixed temporal order, this Hebbian learning will eventually lead to similar mnemonic representations (i.e. patterns of firing rates) for the two stimuli. Thus, an associative memory will be formed.

According to this view, the tendency of neurons to have sustained activity for sets of neighboring stimuli presented in a fixed temporal sequence is a manifestation of this association at the single neuron level. But this association can be only be formed if the memory trace following one stimulus is maintained across the inter-trial-interval (ITI). Theoretical considerations predict that the sustained activity following a specific stimulus, will be evident during the ITI as in the inter-stimuli interval (ISI), because the activity evolves automatically, in a mechanical fashion, irrespective of the behavioral relevance of the stimulus. Since in the DMS task, in half the trials the sample and test stimuli are identical, this propagation of the stimulus selective activity during the ITI could serve as a vehicle to transmit information about the temporal order of the sample stimuli.

We report here that IT neurons indeed have a stimulus selective sustained activity during the ITI. We use a simulation of a large network of (integrate and fire) neurons, to illustrate the development of this sequence of events. In this simulation, the sustained activity during the ITI generates the temporal correlations in the delay activity similar to the findings of Miyashita [1].

Results

Figure 1 illustrates the sequence of events and stimuli used in this study. We recorded the activity of 314 visually responsive cells in IT cortex. Twenty three neurons (7.3%) showed stimulus selective delay activity. Figure 2(A,B) shows examples of two neurons with stimulus selective delay activity. Responses are shown to three different stimuli when presented as the sample stimulus, and when presented as the test stimulus. Note that the rasters and histograms on the left and right halves of the figure are not temporally contiguous. This is because we were interested in the neuronal activity elicited by a specific stimulus, when presented as a test stimulus, irrespective of whether it matched or did not match the sample stimulus.

Figure 2A demonstrates responses from an IT neuron which had highly selective delay activity. In this case stimulus #14 (shown in Figure 1B, top row, second from left) elicited the most vigorous firing during its presentation and following it. Note that the delay activity is evident following the test stimulus, as after the sample stimulus. Furthermore, the delay activity evoked by stimulus #14 following the test stimulus survived throughout the ITI, which lasted 6 - 7 sec, and was evident until the presentation of the next sample stimulus in the following trial (Figure 2A, right column). The few trials in which activity could be seen in the pre-sample intervals (leftmost column) are all cases when the test stimulus in the previous trial was stimulus #14 (rightmost column, marked by grey background). Thus, the conventionally defined "spontaneous" activity is affected by the identity of the last stimulus seen. We therefore analyze the activity prior to the sample stimulus according to the identity of the previous test stimulus (rightmost column, Figure 2A and B). Thus, in Figure 2B we present only the pre-stimulus activity before the next trial. This cell has a more widely distributed selective delay activity. The delay activity following a specific test stimulus is

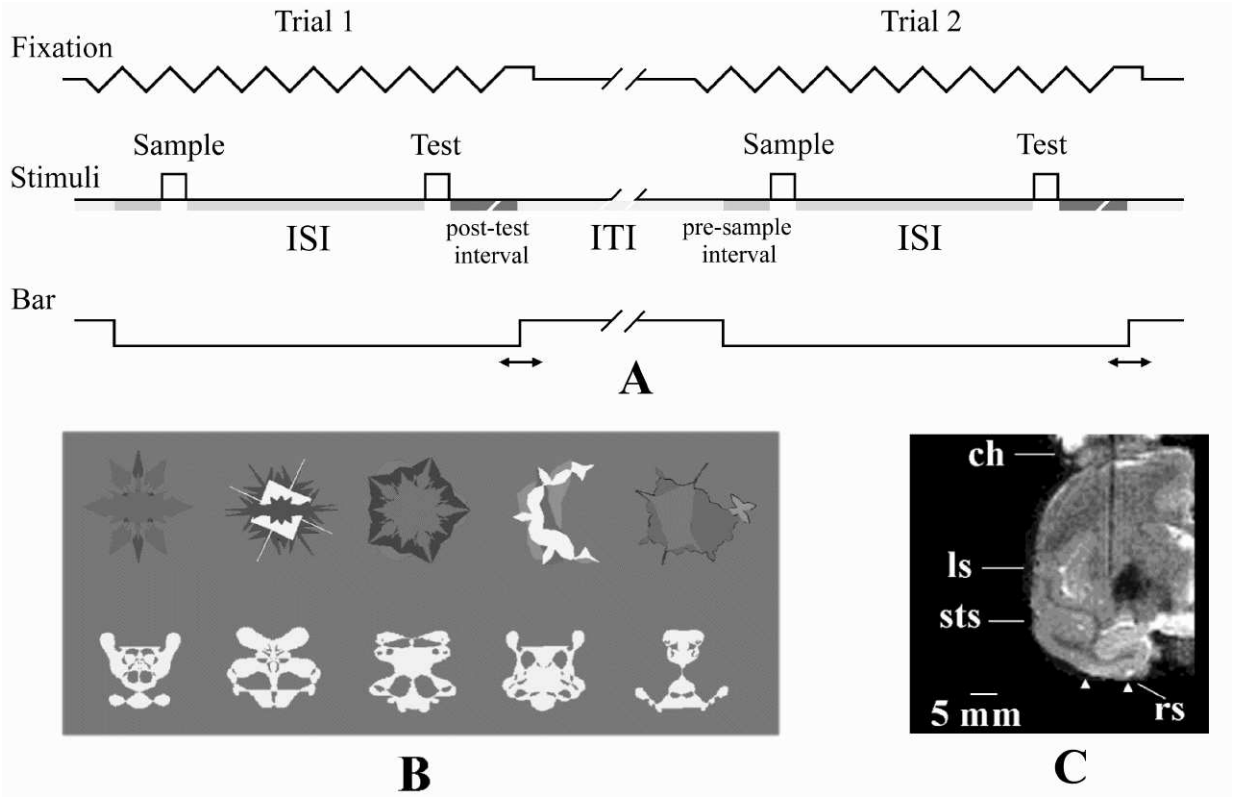


Figure 1: A: Schematic sequence of events in the DMS task: a trial began with the presentation of a flickering dot (at 1 Hz) at the center of a PC monitor screen. The monkey was required to press a lever in response to the flicker onset. Bar press led to the presentation of a first, sample stimulus 1 sec later. The second, test stimulus, was presented after a fixed ISI (usually 5 sec). The test stimulus matched the sample stimulus in half the trials. Both stimuli were presented at the center of the screen. After a variable post-test stimulus interval (500-1500 msec) the central dot stopped flickering and turned bright. This served as a GO signal for the monkey to shift the bar (two sided arrow) left if the test stimulus matched the sample stimulus, and right if the two were different, and then to release the lever to get a fruit juice reward. Both monkeys performed the task above 90% correct. A set of 30 color stimuli were presented in a fixed temporal order during the training session. B: Examples of stimuli used. Top row: fractals, bottom row: Fourier descriptors. C: A coronal MRI image of the right hemisphere of the brain of one of the studied monkeys. Abbreviations used: ch: recording chamber. ls: lateral sulcus. sts: superior temporal sulcus. rs: rhinal sulcus. During the imaging a tungsten electrode was placed at the center of the recording chamber (dark vertical shadow). Its depth corresponds to the approximate location of the tip of the guide tube during recording. The area explored during the experiments between the rhinal sulcus and anterior-medial-temporal sulcus is marked by triangles.

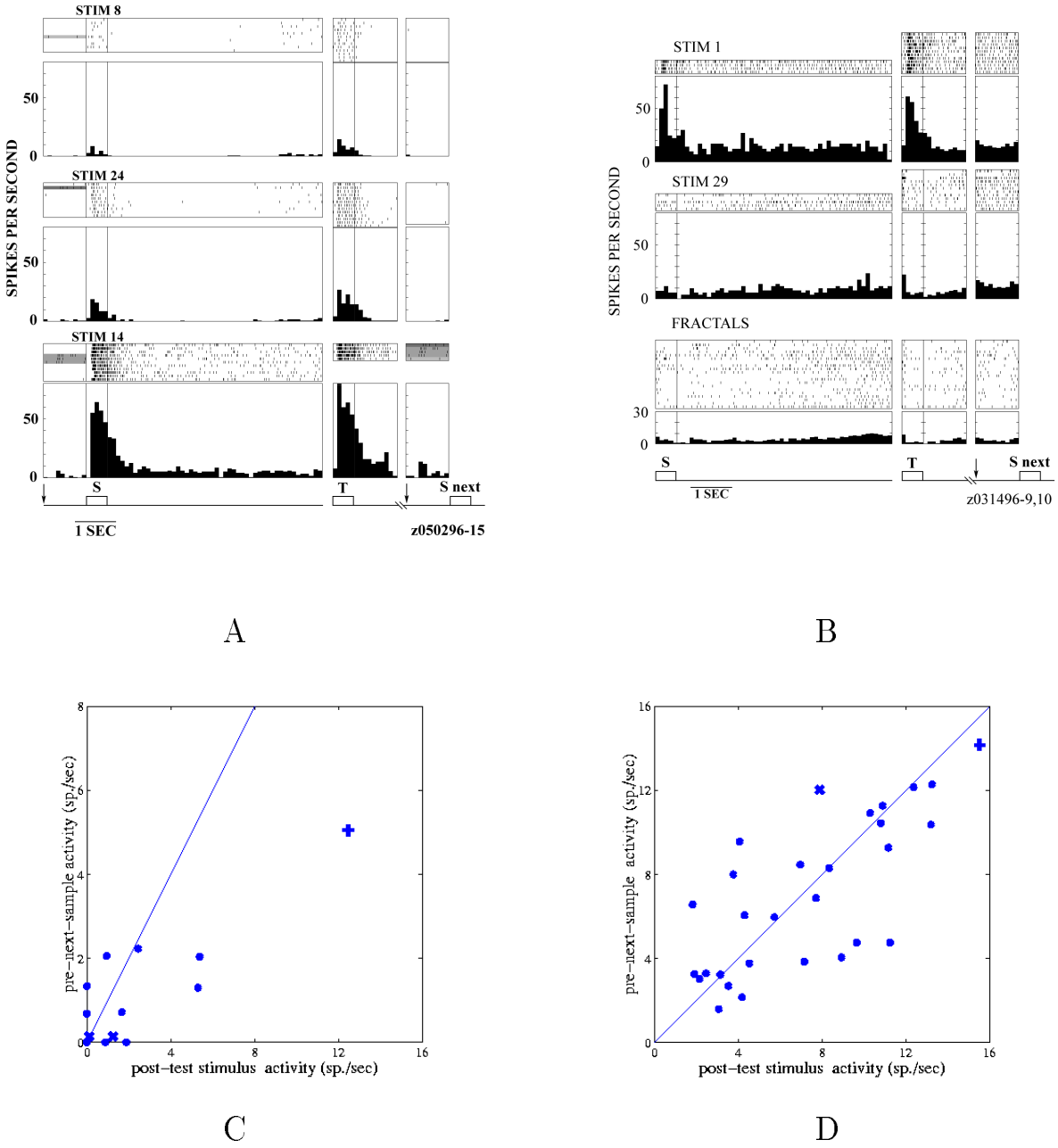


Figure 2: A and B: Example of two neurons that showed sustained activity throughout the ITI. Abbreviations used: Arrow down: bar press, S: sample stimulus, T: test stimulus, S next: sample stimulus in the following trial. Note that the different intervals within the trial are sorted and organized according to the identity of the corresponding stimulus. Consequently, the number and order of the rasters for the sample and test stimuli are not in register: the data shown during and following the test stimulus are combined for the same test stimulus across match and non-match conditions. A: a neuron with highly selective delay activity. The right arrow down corresponds to a bar press in the beginning of the next trial, 6-7 sec after the termination of the trial (indicted by broken lines). Note that stimulus selective delay activity is as clear following the test as following the sample stimulus (#14). Sustained activity following the test stimulus was evident throughout the ITI, until the onset of the sample stimulus of the next trial. B: An example of a neuron with more widely distributed selective delay activity. The sustained activity following a specific test stimulus is maintained throughout the ITI until the presentation of the next sample stimulus (see stimuli #1 and #29 compared to all the fractal stimuli, which don't elicit a delay activity). C and D: Scatter plots of the average delay activity in the last period of the ITI (one second before the sample stimulus of the next trial), as a function of the average sustained activity in the beginning of the ITI (following the test stimulus). Scatter plots C and D correspond to the data from the neurons shown in A and B, respectively. Each data point is an average across different presentations of a given test stimulus. (Numerous stimuli did not elicit any activity in the cell depicted in A, and therefore the data point in the origin (0,0) of Figure 2C represents multiple stimuli). The + sign denotes the response to the best stimulus (#14 in C, #1 in D) while the X symbols depict the response to the ineffective stimuli (#24 and #8 in C) and less effective stimulus (#29) in D. The diagonal lines indicate points of equal response in the two time epochs.

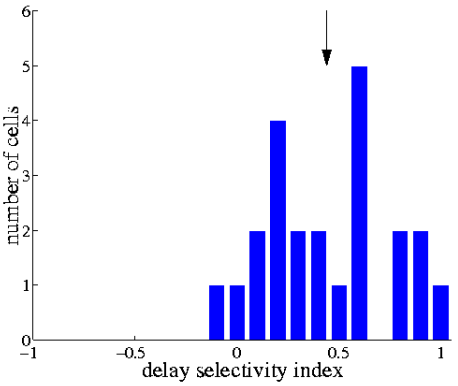
maintained throughout the ITI until the presentation of the next sample stimulus (see stimulus #1 in Figure 2B). The scatter plots in Figures 2C,D show the delay activity in the last period of the ITI (pre-next-sample stimulus activity) as a function of the delay activity in the first interval of the ITI (the post-test stimulus activity) for the neurons shown in Figure 2A and B, respectively. Each datum is an average across all trials with the same test stimulus.

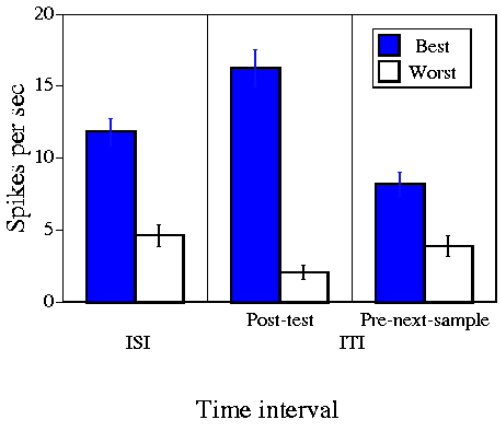
To evaluate the reliability of transmission of information across the ITI in the population of delay activity neurons, we compute for each neuron a delay selectivity index: We define the "best" and "worst" stimuli respectively as the ones that elicit the strongest and weakest activity during the beginning of the ITI (the post-test stimulus period). The delay selectivity index is defined as $(R_{best} - R_{worst})/(R_{best} + R_{worst})$ where R is the activity during the last period of the ITI (pre-next-sample stimulus activity). The delay selectivity index is bounded between the values of [-1,1]. A value of zero indicates that in the end of the ITI there is no difference between the responses to the stimuli that elicited a very different response during the beginning of the ITI (i.e. no propagation of information). More positive values indicate the maintenance of the differential response across the ITI. Note that by definition, the "best" and "worst" stimuli will elicit a different response in the first period of the ITI (because this is the basis of the selection of the stimuli) even if the neuron's delay activity is not truly stimulus selective (i.e. if the difference between the "best" and "worst" stimuli is due only to random fluctuations in the response). The crucial point is whether this differential response is maintained throughout the ITI and evident in the last period of the ITI.

Histograms depicting the distribution of the delay selectivity index across the population of the 23 neurons with selective sustained activity are shown in Figure 3A. The sustained activity tended to survive through the ITI. The average value of the delay selectivity index for this group of neurons was 0.44, (similar to the level of selectivity when the same measure was applied to the ISI, revealing a delay selectivity index of 0.52). This index was significantly different from zero, (one group t-test, $P < 0.0001$). The average activity following the "best" and "worst" stimulus for the different time intervals is shown in figure 3B. The difference in response is evident in the ISI, in the "classical" delay period. This difference was also highly significant in the last period of the ITI, before the presentation of the next sample stimulus (paired t-test, $P < 0.0001$). Finally, a cell-by-cell analysis of the activity in the initial and final periods of the ISI and the ITI, shown in Fig 3C, demonstrates that in the vast majority of neurons, the difference in response between the "best" and "worst" stimulus was maintained across the ITI. In fact the magnitude of the differential response in the end of the ITI was almost identical the corresponding one at the end of the ISI (5.0 vs 4.6 spikes/sec, correspondingly).

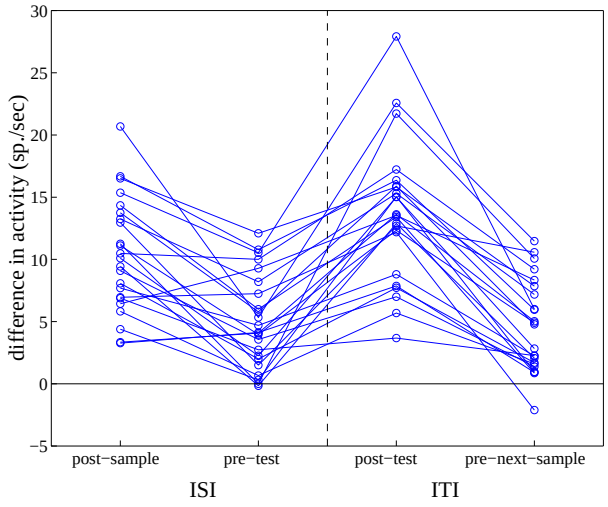
The sustained activity following the best sample stimulus was disrupted if the test stimulus was different from the sample, in accordance with previous findings [10]. Thus, the sustained activity depended on the identity of the last stimulus seen, be it sample or test stimulus.

There was a strong positive correlation between the visual response to the sample stimulus and the delay activity in the following ISI, when the *average* activity for each stimulus was considered (average Pearson r = 0.69, N=23). But the correlation between the activity during the presentation of the "best" sample stimulus and the ISI delay activity on a *trial by trial basis* was much weaker. In fact, the visual response and the activity in the last second of the ISI were generally uncorrelated (average Pearson r = 0.10). During the post-test period the difference in sustained activity (between the "best" and "worst" stimulus) was greater then during the corresponding period in the ISI (13.66 vs 10.34 sp/sec, correspondingly), but this difference was not evident at the end of the two intervals (5.0 vs 4.6 sp/sec, correspondingly). The response in the post-test period was also usually somewhat attenuated when the test stimulus matched the sample stimulus (13.92 vs 17.71 sp/sec for the "best" stimulus in the match vs. non-match conditions), but again it was not significantly different at the end of the ITI (7.80 vs 8.05 sp/sec, correspondingly). In summary, although the delay activity immediately following a specific stimulus was dependent on





A B



C

Figure 3: A: A histogram depicting the distribution of the delay selectivity index across the population of selective sustained activity neurons (N=23). The index is the difference in activity in the end of the ITI, following the "best" and "worst" stimuli, divided by the sum of the two responses. The "best" and "worst" stimuli were chosen according to the activity they evoked during the beginning of the ITI (the post-test period). Maintained selectivity throughout the ITI, is manifest in positive values. The average sustained selectivity index was 0.44 (indicated by the arrow). This corresponds to a response more than twofold stronger for the "best" stimulus compared to the "worst" stimulus in the end of the ITI. B. Average neuronal activity for the "best" and "worst" stimuli during the ISI, and the beginning (post-test) and end (pre-next-sample) periods of the ITI. Error bars indicate the SEM. Note that the neuronal activity during the ISI was based on the trials in which the "best" and "worst" stimuli appeared as sample stimuli, while the activity in the ITI was according to the identity of the test stimulus. C. The difference between the activity elicited by the "best" and "worst" stimuli during the corresponding initial and last periods of the ISI and ITI, shown individually for each neuron. The vast majority of the neurons maintain their differential response during the ITI, as in the ISI

the magnitude of the visual response that could be vary from trial to trial for different reasons, the final level of delay activity (a few seconds later) was constant. All these pieces of evidence are in line with the suggestion that the delay activity is a result of the neural network properties, rather than a change in the state of the single neuron alone, triggered by the visual response (see also [9] and below).

Is there a functional role to the propagation of delay activity across the ITI? We suggest that it may serve to allow the generation of sustained activity for neighboring stimuli that are repeatedly shown in a fixed temporal order. Indeed, we observed that the few stimuli that evoked sustained activity were often neighboring stimuli, as has been reported earlier by Miyashita [1]. An example of such a neuronal response is shown in Figure 4. Clustering of delay activity according to the serial position number (SPN) of the stimulus is obvious both in the ISI and in the two ends of the ITI.

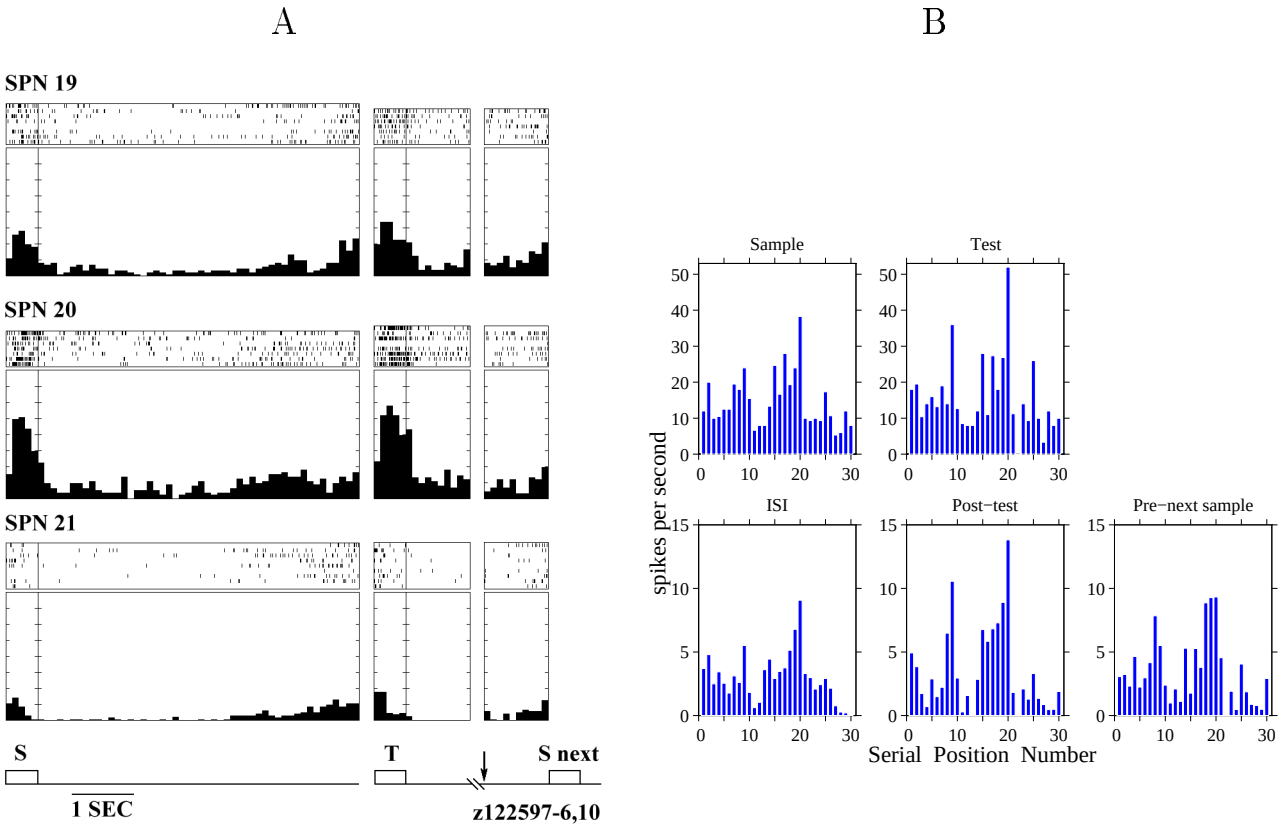


Figure 4: A: Raster displays and spike density histograms of one neuron for three consecutive stimuli of the thirty stimuli that were presented in a fixed temporal order during the training stage. SPN denotes the serial position number of the stimulus in the training sequence. All other aspects of the data are presented in the same way as in Figure 2B. B: A histogram depicting the response of the same neuron during the presentation of the sample and test stimuli (top) as well as the sustained activity (bottom) during the ISI and the two ends of the ITI (post-test-stimulus period, and pre-next-sample period) as a function of the SPN. Similar clustering of the sustained activity according to the SPN is observed in all time epochs.

Such a context dependent associative memory is formed in three stages according to the attractor picture: first, the uncorrelated (context independent) attractors build up. In the next phase the information about the patterns of activities can be propagated from trial to trial by the sustained activity in the ITI. This activity leads to the buildup of correlations between stimuli from consecutive trials; finally the pattern of connectivity is such that the representation of every stimulus reflects the temporal context: delay activity distributions corresponding to neighboring stimuli in

the training sequence are more correlated than the ones corresponding to distant stimuli in the training sequence. The ITI selective activity is an essential building block for the detection and memorization of temporal correlations in the statistics of the flow of stimuli and it is sufficient to correlate not only the nearest neighbors but also stimuli that are up further apart in a sequence.

In the next section, following [7, 11] we exemplify the mechanism underlying the formation of the temporal correlations by taking three snapshots of the behavior of the modeled network corresponding to the three stages.

Model Neural Network

We present a model neural network to illustrate how such a context dependent associative memory can be formed. The model is a direct implementation of the one proposed by N. Brunel [11] (for more details see the Methods section). Here we focus on the role of the ITI selective activity and we therefore expanded the analysis of the dynamics and show the typical behavior of the model neurons during all the learning stages, as they would appear in cortical recording in each interval of the trial (visual response, ISI and ITI).

The network is composed of excitatory and inhibitory neurons, represented (for simplicity) by afferent currents and output rates. Each neuron in the network receives three types of input: from recurrent excitatory connections from other neurons in the same module; non-selective, excitatory afferents from other areas of cortex, and local, non-selective inhibitory afferents. The statistics of the input currents determines the firing rates as in [12]. The excitatory neurons in the module belong to sub-populations, each responding (for simplicity) to only one stimulus.

Figure 5 shows the development of delay activity in model neurons during the training process using stimuli that were repeatedly presented to the network in a protocol identical to the one we used in the Results section, above. During stimulation (when the sample or test stimuli are shown), an extra current is injected in the sub-population of neurons responding to the stimulus presented. The elevated activity of these neurons leads to an increase in the activity of the population of inhibitory neurons, which always reflects the global activation of the excitatory population. As a result, the activity of the other sub-populations that are not activated by the stimulus is depressed.

In the first stage, the strength of the inter-class (between sub-populations responding to different stimuli) and intra-class (within sub-populations) connections is randomly chosen and each neuron shows a stimulus selective visual response but no delay activity. The joint firing of two neurons activated by the same specific stimulus (for instance: neuron A, and a similar neuron from the same group of neurons responding to stimulus #2) leads to the potentiation of the connection between the two. Analogously, the inter-class connections tend to be depressed. With enough repetitions of the same stimulus, there are enough potentiated synapses that the network can sustain enhanced activity even after the evoking stimulus has been removed: each neuron in the subpopulation excites the others through the potentiated synapses. At the end of this stage, delay activity distributions (attractors) are formed for each specific stimulus. This network property appears suddenly and is observed as a stimulus specific delay activity, (shown in stage 2).

Because of this stimulus specificity one should find a positive correlation between the visual response and the following delay activity. This is in correspondence with what we found (see also [4]). On the other hand, the sustained activity evoked by one specific stimulus is not affected by fluctuations in strength of the visual response from *trial to trial*, as we report above. This is because the delay activity is triggered by the visual response, but it is sustained by the pattern of activity of all the neurons in the same module. The visual response determines the initial condition: all the stimuli which evoke patterns of activity that are in the same basin of attraction lead to the same final steady state (attractor), irrespective of the fluctuations of individual neurons' activity (see also [9]).

In Stage 2, stimulus selective delay activity exists, but the patterns of delay activity across the population of neurons are initially not overlapping (i.e. each neuron has sustained activity to only

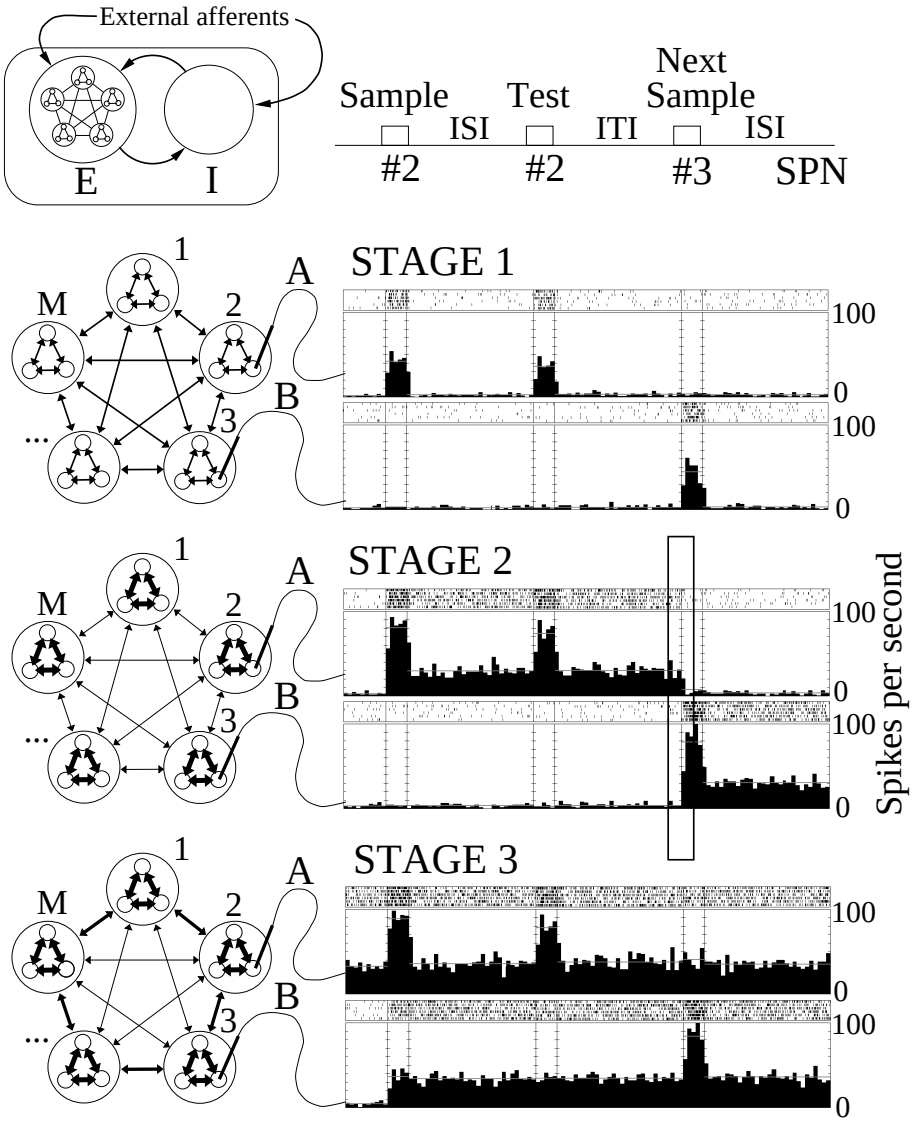


Figure 5: Three snapshots of the behavior of the modeled network corresponding to the three stages of the development of associative memory. The left column depicts the scheme of the model network in the three stages. A network module is composed of two interconnected populations: stimulus selective excitatory neurons and inhibitory non-selective neurons. Both populations receive external excitatory afferents from other modules in the cortex (inset in top left panel). In the excitatory population (shown in left panel, below inset) , each circle denotes a sub-population of neurons selective to a specific stimulus. Populations corresponding to stimuli that are nearest neighbors in the training sequence of length M are arranged in the picture so that they are near in the circular chain, for the convenience of presentation. Only five sub-populations are represented in the picture. The arrows denote the synaptic connections. There are essentially two classes of connections: intra-class synapses connecting neurons responding to one specific stimulus; inter-class synapses, connecting neurons responsive to different stimuli. Thicker arrows denote a higher number of potentiated connections. Two neurons (denoted by A, and B) are monitored during trials of the DMS task in which stimuli #2 and #3 (numbered according to their serial presentation order (SPN)) are repeatedly presented. (for illustration purposes only, we present 10 trials in which the sample and test stimulus (#2) are the same). Their neuronal activity in the different stages is shown in the right column (the dotted gray lines represent the average activity in each interval). See the text for an explanation of the dynamics.

one stimulus). The delay activity of a specific sub-population is triggered by the presentation of the corresponding stimulus and ends with the presentation of a different stimulus. This is because the global inhibition generated by a different visual stimulus is enough to suppress the activity of this sub-population, to its spontaneous activity.

When the test stimulus matches the sample, neuron A will have delay activity following the test stimulus (#2), until the presentation of the next stimulus, which generates a visual response in all neurons of sub-population #3 (including neuron B). This joint activity (within a time window of less than 100 msec, at the end of the ITI of stimulus #2) allows for Hebbian strengthening of the synapse between the two neurons. The result of this unsupervised learning is that in the final stage (Stage 3) the neurons show sustained activity also for the neighboring stimuli of the eliciting stimulus. This is the phenomenon that was reported above (Figure 4; see also [1]). The spreading of the delay activity to the neighboring stimuli is limited to a maximal distance of a few stimuli (in our case the parameters are such that the maximal distance is 5. This limitation is not obvious in the figure because we show only the immediate neighboring stimuli in the sequence). This is because the inhibition is faster and stronger than excitation. Thus, the dampening effect of the total inhibition becomes dominant, whenever the total excitation tends to grow. Moreover, potentation is dependent on the joint level of activity of the two neurons. The delay activity is generally weaker than the visual activity to a given stimulus. Therefore, the delay activity of neuron A elicited by the neighboring stimulus (#3) will usually be weaker than the activity evoked by the original stimulus (#2), and the chain reaction will be limited. (for more details: see [7, 13, 11]).

Discussion

The most prominent and novel finding reported here is that stimulus selective delay activity in IT cortex persists across the ITI. Recently, an analogous type of sustained activity which persisted across the ITI was reported in PF cortex [14]. This activity was not related to eye-movements, and was seen also in a monkey that was never trained on a memory task, indicating that it evolves automatically. The sub-area within PF in which these face-selective neurons are found, receives strong input from IT.

Stimulus selective delay activity was considered to encode the memory trace during the ISI [2, 1, 15]. Recently, it was shown that sample-specific delay activity in PF cortex is maintained throughout the trial, even when intervening stimuli were presented, whereas delay activity following the sample stimulus was disrupted by intervening stimuli in IT cortex [10, 16]. These authors concluded that PF cortex may subserve "active" working memory, whereas IT cortex, may contribute to an automatic detection of stimulus repetition. Our results are in agreement with the hypothesis of a "passive", automatic memory in IT.

One could possibly suggest that the sustained activity following the test stimulus was due to active working memory, since the identity (match/non match) of the test stimulus must be remembered for executing a correct response. However, the stimulus specific sustained activity following the test stimulus, was evident even after the reward, when memory of the stimulus was no longer required. The delay activity also cannot serve as the mnemonic trace of the sample stimulus throughout the trial, since it was disrupted by the presentation of a different test stimulus. Thus, the sustained activity seems to reflect the last familiar stimulus seen irrespective of its relevance to the behavioral task.

Could the sustained activity be a result of unmonitored eye movements? The sustained activity was stimulus specific and reproducible. It occurred both in the ISI, when the animal gazed at the center of screen, as was observed by the video camera, and in the ITI, when the monkey was clearly seen making eye movements. Thus, it is highly unlikely that the delay activity was caused by systematic eye movements following a specific stimulus. Furthermore, previous reports have shown that eye movements do not influence the delay activity of IT neurons in the ISI [4].

We suggest that the propagation of activity across the ITI may serve to generate the synaptic structure required to form correlations between the mnemonic representations (delay activity distributions) of successive stimuli in a sequential training protocol [9].

In what circumstances would such a mechanism have behaviorally observed consequences? It was suggested in [11] that it would be highly effective in a paired association task, in which retrieval from long-term memory of the pair member associated with a given cue is required. Such associations are formed by repeated presentations of the paired associates, and monkeys with lesions of the temporal lobe (rhinal cortex), or the connections between inferotemporal and prefrontal cortex show marked impairment in this task [17, 18]. Recently, Sakai and Miyashita [3] have trained monkeys to perform such a paired association task. They have demonstrated that neurons in IT cortex, selectively responded to both pictures of the paired associates. Furthermore, the example they present clearly shows that the neuron had a similar level of delay activity following the presentation of either of the two paired stimuli as a cue, suggesting that both stimuli now evoke the same pattern of delay activity, i.e. the same attractor.

A key requirement for the buildup of an attractor network is that neurons are organized in local groups with similar stimulus specificity, or that neurons with similar specificity are preferentially connected. Indeed, Tanaka and colleagues [19] have clearly established that neighboring neurons in inferotemporal cortex tend to have similar stimulus preferences, using both standard microelectrode and optical imaging techniques.

Lately a similar model of attractor dynamics was suggested for the generation of invariant face and object recognition in vision. In essence it suggests that cells in IT cortex, become invariant to the viewing angle by taking advantage of the fact that usually faces or objects are seen from

different views in a temporal sequence as one is manipulating an object or moving in space [20, 21].

We conclude that the stimulus selective sustained activity in IT reflects a "passive", automatic memory. The persistence of stimulus selective activity across the ITI may serve as the necessary link to generate associations between neighboring stimuli. This may be accomplished by a modification of the synaptic structure, so that correlations between the neural representations of successive stimuli are formed.

Such a scheme of association may be relevant for navigation in an environment we are not well acquainted with. In such circumstances we usually remember the specific route we have taken, rather than rely on a cognitive map of the environment. Navigation in such circumstances relies heavily on remembered snapshots of visual scenes from specific angles, with one cue leading us to the next expected landmark. Interestingly, lesions in parietal cortex typically lead to a failure in grasping the spatial relationships between places, (i.e. a failure to generate a cognitive map), with intact landmark recognition [22]. Temporal lobe lesions in humans, on the other hand, often result in topographical disorientation in novel environments, when landmarks along the route are used [23]. Furthermore, such topographical agnosia often co-occurs with prosopagnosia (inability to recognize familiar faces) [24, 25]. This paradoxical finding is better understood if attractor dynamics in the temporal lobe is the common neural mechanism underlying the two mnemonic functions.

Methods

Behavioral task and Visual stimuli. The activity of single neurons was recorded from IT cortex while monkeys performed a visual DMS task. The monkeys were seated in an isolated experimental chamber with a background illumination of 2 cd/m^2. The only objects in front of the monkey were the PC monitor and a video camera. The background luminance of the screen was 12 cd/m^2, while the colored images were high contrast pictures. A set of 30 color stimuli were presented in a fixed temporal order during the training session. Fifteen were fractal stimuli, and the rest were Fourier descriptors.

Animals and surgical procedures. Two rhesus monkeys (Macaca mulatta) weighting 6 -7 kg were used. A head post and a recording chamber were implanted above anterior-ventral IT cortex under general anesthesia with nembutal (25-30 mg/kg). The monkeys were given antibiotics and analgesics postoperatively, and were allowed sufficient time for recovery after surgery. All experiments, MRI tests and surgical preparations were performed in accordance with NIH and Hebrew University guidelines for use of laboratory animals for experiments.

Anatomical MRI. We applied magnetic resonance imaging (MRI) technique using a Biospec 47/40 device (Bruker) to verify the position of the recording chamber, relative to the area of interest. A series of coronal T2-weighted images (13-15 consecutive 2 mm slices) were recorded covering the whole area of interest in the monkey brain. A tungsten electrode (diameter = 200 microns) was inserted through the chamber center above the area explored during the actual recording sessions (Figure 1C). The area explored was between the rhinal sulcus and anterior-medial-temporal sulcus. The images were recorded using a spin-echo sequence with the following parameters: field-of view of 13x13 cm, 256x256 data matrix, a RARE factor of 8, TR/TE of 3000/23 ms and 8 scans yielding an effective T2-weighted contrast images corresponding to normal spin-echo taken with TE=70 ms. The monkeys were anesthesised during the imaging session which lasted about 15 min.

Recording technique & Data analysis. Single unit activity was monitored in four hemispheres of two monkeys using standard recording techniques. Due to technical limitations (data transfer between computers, generation of new stimuli, etc), neuronal activity was registered during the period between the beginning of the trial (presentation of flickering dot) and bar release. Therefore, the activity during the ITI was monitored in two discrete periods: 1) The post-test stimulus activity, i.e. the firing rate between the test stimulus offset and the bar release. 2) The pre-next-sample activity, defined as the firing rate in the interval prior to the next sample stimulus,

from the presentation of the flickering dot to the sample stimulus onset. The activity during the ISI was defined as the firing rate in the interval between the sample and test stimuli. The first 200 msec following stimulus offset in the ISI and ITI were excluded to avoid the effects of a possible visual response. Neurons were considered to have a stimulus specific delay activity if the firing rates for the various stimuli during both the ISI and post-test stimulus period were statistically different using one way analysis of variance (ANOVA, $p<0.001$).

Details of the model. The parameters are the same as in [11]. The statistics of the input currents determines the firing rates as in [12], where the current-to-rate transduction function was calculated for leaky integrate-and-fire neurons. The integration time constant for excitatory (inhibitory) neurons is 10 ms (2 ms) and the emission threshold is $20mV$ above the resting level. Each neuron receives 10^4 afferents from randomly selected excitatory neurons of the same module, 2×10^3 afferents from the population of inhibitory neurons and an external current from other unspecified areas. The mean synaptic efficacies are chosen in such a way that in the first stage, when the synaptic matrix is still not structured, the average spontaneous activity is 3.0 sp/s for the excitatory neurons and 4.1 sp/s for the inhibitory neurons (the EPSPs are: $J_{E\to E} = 0.035$mV, $J_{E\to I} = 0.054$mV, $J_{I\to E} = J_{I\to I} = -0.141$mV). The external mean excitatory current is the same as the mean recurrent excitatory current when all the neurons have spontaneous activity. During stimulation an extra gaussian current is injected in the neurons of the sub-population (fraction $f = 0.01$ of the excitatory neurons in the network) corresponding to the activated stimulus ($\mu = 8.25$ mV/ms, $\sigma^2 = 0.9$ mV^2/ms).

Only the excitatory synapses in a module are modifiable and each synapse has two potentiation levels [26]. The high level (potentiated state) corresponds to a synaptic efficacy which is 4.4 times larger than the low level (depressed state). Synaptic transitions between the two levels depend on the mean rates of the pre and post-synaptic neurons: LTP (long term potentiation) corresponds to the transition between the low level and the high level and occurs with probability $p_+ = 0.2$ if the pre and the post-synaptic neurons are simultaneously activated by the stimulus (i.e. following each repetition, a mean fraction p_+ of the depressed synapses that are connecting active neurons makes a transition to the potentiated state). If one neuron is activated by a stimulus (high rate) and the other carries selective delay activity elicited by the previous stimulus seen, then potentiation occurs with probability $p'_+ = ap_+$ (a=0.015). LTD (long term depression) occurs with a probability $p_- = 0.2$ when one neuron is activated by the stimulus while the other is at spontaneous rate.

Acknowledgments

We wish to thank Gadi Goelman for technical support in the MRI testing, Michail Dvorkin for development of some of the data analysis tools, and Nicolas Brunel for his generous help in reproducing his simulations, and to Robert Shapley that pointed out the possible relevance of the connection between prosopagnosia and navigation problems in patients with temporal lobe lesions. Daniel Amit and Shaul Hochstein offered insightful comments on earlier versions of the manuscript. Supported by grants from the Israel Academy of Science, Israel National Institute of Psychobiology (V.Y.) and a McDonnel-Pew grant for cognitive neuroscience (E.Z).

References

[1] Miyashita, Y. (1988) Neuronal correlate of visual associate long-term memory in the primate temporal cortex. *Nature* **335**, 817-820.

[2] Fuster, J. M. and Jervey, J. P. (1981) Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**, 952-955.

[3] Sakai, K. and Miyashita, Y. (1991) Neural organization for the long-term memory of paired associates. *Nature* **354**, 152-155.

[4] Nakamura, K. and Kubota, K. (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J. Neurophys.* **74**, 162-178.

[5] Wilson, F.A.W., Scalaidhe, S.P.O., and Goldman-Rakic, P.S. (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex *Science* **260**, 1955-1958.

[6] Rao, S.C., Rainer, G., and Miller, E.K. (1997) Integration of what and where in the primate prefrontal cortex *Science* **276**, 821-824.

[7] Amit, D.J., Brunel, N., and Tsodyks, M.V. (1994) Correlations of cortical hebbian reverberations: Theory versus experiment. *J. Neuroscience* **14**, 6435-6445.

[8] Amit , D.J. (1995) The Hebbain paradigm reintegrated: Local reverberations as internal representations. *Behavioral and Brain Sciences* **18**, 617-657.

[9] Amit, D.J., Fusi, S., and Yakovlev, V. (1997) Paradigmatic working memory (attractor) cell in IT cortex. *Neural Computation* **9**, 1071-1093.

[10] Miller, E.K., Li, L., and Desimone, R. (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* **13**, 1460-1478.

[11] Brunel, N. (1996) Hebbian learning of context in recurrent neural networks. *Neural Computation* **8**, 1677-1710.

[12] Amit, D. J. and Brunel, N. (1997) Global spontaneous activity and local structured (learned) delay activity in cortex. *Cerebral Cortex* **7**, 237-252.

[13] Griniasti, M., Tsodyks, M. V., and Amit, D. J. (1993) Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Computation* **5**, 1-9.

[14] Scalaidhe, S.P.O., Wilson, F.A.W., and Goldman-Rakic, P.S. (1997) Areal segregation of face-processing neurons in prefrontal cortex *Science* **278**, 1135-1138.

[15] Fuster, J. M. (1990) Inferotemporal units in selective visual attention and short-term memory. *J. Neurophys.* **64**, 681-697.

[16] Miller, E. K., Erickson, C., and Desimone, R. (1996) Neural mechanisms of working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154-5167.

[17] Murray, E.A., Gaffan, D. and Mishkin, M. (1993) Neural substrates of visual stimulus-stimulus association in rhesus monkeys.*J. Neurosci.* **13**, 4549-4561.

[18] Gutnikov, S.A., Yuan-Ye, M., and Gaffan, D.(1997) Temporo-frontal disconnection impairs visual-visual paired association learning but not configural learning in macaca monkeys.*Euro. J. Neurosci.* **9**, 1524-1529.

[19] Tanaka, K. (1996) Inferotemporal cortex and object vision. *Ann. Rev. Neurosci.* **19**, 109-140.

[20] Wallis, G. and Rolls, E. (1997) Invariant face and object recognition in the visual system. *Progress Neurobiol* **51**, 167-194.

[21] Bartlett M. and Sejnowsky T. (1998) Learning viewpoint invariant representations from visual experience using attractor networks. *Network.* **in press**.

[22] DeRenzi, E., Faglioni, P. and Villa, P. (1977) Topographical amnesia.*J Neurol Neurosurg Psychiatry* **40**, 498-505.

[23] Whiteley, A. M. and Warrington, E. K. (1978) Selective impairment of topographical memory: a single case study. *J Neurol Neurosurg Psychiatry* **41**, 575-578.

[24] Landis, T. Cummings, J. L. Benson, D. F. and Palmer, E. P. (1986) Loss of topographic familiarity. *Arch Neurol* **43**, 132-136.

[25] Maguire, E. A., Burke, T. Philips, J. and Staunton H. (1996) Topographical disorientation following unilateral temporal lobe lesions in humans. *Neuropsychologia* **34**, 994-1004.

[26] Petersen, C. C. H., Malenka, R. C., Nicoll,R. A. and Hopfield, J. J. (1998) All-or-none potentiation at CA3-CA1 synapses, *Proc. Natl. Acad. Sci.* **95**, 4732-4737.