IEEE International Conference on Neural Networks, 1997, I-372-376.

A Neural Network Model of Hippocampally Mediated Trace Conditioning

William B Levy and Per B. Sederberg
Department of Neurological Surgery
University of Virginia Health Sciences Center
Charlottesville, Virginia 22908, USA
E-mail: wbl@virginia.edu, pbs5u@virginia.edu

Abstract

In this paper a simple biological model of hippocampal region CA3 simulates the learning of hippocampally dependent trace classical conditioning. In this biologically based model, the time span of the associative modification rule is 5-fold less than the trace interval, implying that recurrent cell firing must play a significant role in encoding the trace interval. The results show that this simple network— with its moderate time spanning synaptic modification rule and sparse connectivity— can learn to span trace intervals comparable to those in rabbit eyeblink experiments [10, 11]. That is, the model learns to produce a cell firing pattern equivalent to an anticipatory unconditioned stiumulus. This anticipatory pattern contains the information needed to intercept the unconditioned stimulus with a conditioned response because it is delivered at an appropriate time before the actual unconditioned stimulus.

1. Introduction

Although the cerebellum is the identified locus for learning and remembering classical conditioning [12], an intact hippocampus is also necessary for a mammal to learn the form of classical conditioning called trace conditioning [10, 11]. In the trace paradigm, the animal must learn an interval of time called the trace which spans the temporal gap between the offset of the conditional stimulus (CS) and the onset of the unconditioned stimulus (UCS). During this period, the animal must withhold the conditioned response (CR) until just before the UCS. In the classical conditioning paradigms that use an air puff to the eye as a UCS, this requirement of trace conditioning means that the eyeblink CR is produced by the animal in a way that will block the air puff UCS from hitting the eye. Thus, the animal must correctly time the eyeblink to just anticipate the UCS. Without a hippocampus during learning, the eyeblink CR is either not learned or is delivered at the wrong time, i.e. too soon [10, 11].

The trace conditioning paradigm is arguably the simplest sequence learning problem for which an animal needs its hippocampus. The functional mechanism (i.e., the anatomy and physiology) for learning trace conditioning is the problem that we address here by using a minimal biological model.

Learning the trace interval can not be just a function of known associative synaptic modifications in hippocampus. That is, the synaptic modifications that provide the basis for learning across time are too short, by themselves, to explain learning of the trace interval. Synaptic time spanning associations are in the right direction (earlier presynaptic excitation is associated with later postsynaptic excitation) but only span 100 or, at the most, 150ms [8, 3] while the trace interval is usually 250 or 500ms. Previously [8], we proposed that, in addition to the temporal offset of associative synaptic modification, hippocampal circuitry would be required to bridge the trace interval. Here, we demonstrate that a very simple model of hippocampal region CA3 can learn a trace interval longer than the time constant of the associative synaptic modification rule.

2. Methods

The network used in this paper has a simple two-layer architecture. The first layer is the input layer, which represents the entorhinal cortex and dentate gyrus (EC/DG). The input layer projects, one-to-one, onto a sparsely connected layer similar to the CA3 region made up of 1024 neurons. Each neuron in the CA3 layer projects recurrently to 10% of the other cells, similar to the sparse recurrence observed *in vivo*[1, 4]. Neurons are modeled as simultaneously updated McCulloch-Pitts units with a binary output state $z_i(t) \in \{0, 1\}$. Recurrent excitation fires many more neurons than are fired by the EC/DG input.

A competitive scheme determines which cells fire. After computing the busline activation of each neuron,the simulation sorts these activations and sets a variable threshold that

0-7803-4122-8/97 $10.00©1997 IEEE

fires the precise number of neurons for the desired activity level. As part of the competition, a neuron automatically fires whenever its excitatory EC/DG input fires. We performed all simulations at rather low activity levels (5%) analogous to the low firing rates typically observed in the hippocampus.

2.1. The Synaptic Modification Rule

We use an unsupervised Hebbian type postsynaptic associative modification rule, consistent with experimental observations in the hippocampus, to update the synaptic weights [7]. The rule spans time with the addition of a running average of the presynaptic firing state, $\bar{z}_i(t)$. Thus at time t, the synaptic weight $w_{ij}(t)$ is:

$$w_{ij}(t) = w_{ij}(t-1) + \mu z_j(t)[\bar{z}_i(t) - w_{ij}(t-1)], \quad (1)$$

where

$$\bar{z}_i(t) = \epsilon z_i(t) + (1-\epsilon)\bar{z}_i(t-\Delta t), \quad (2)$$

Δt is the simulation time step, and μ is the learning rate.

The multiple time step spanning modfication rule enables us to represent real time in the network. For example for a time step Δt, the decay rate of the running averager $\bar{z}_i$ is:

$$\epsilon = 1 - \frac{\Delta t}{\tau}, \quad (3)$$

where τ represents the time constant of the NMDA receptor, which appears to set the timing-window for associative modification in the hippocampus [2, 8]. In these simulations, we use a network of 1024 cells with $\Delta t = 20ms$(nominal) and $\tau = 100ms$(i.e., $5 * \Delta t$).

2.2. Network Input and Learning

Each simulation has two phases: training (learning) and testing (recall). To reset the network before each presentation of the sequence, neurons are randomly activated to the appropriate activity and allowed to cycle for 10 time steps with synaptic modfication disabled.

During training, the external input drives the network and recurrent synapses modify. The input consists of firing 10 neurons for 8 time steps (160ms), representing the tone CS. The CS is followed by a trace period with no input. The trace period is then followed by firing 15 neurons for 5 time steps (100ms), which represents the UCS air puff (See Figure 1). The network receives the input sequence repeatedly, i.e. enough times to form its own code for the sequence (usually 200 trials). For different simulations the trace periods varied from 5 to 100 time steps (100ms to 2000ms).

After training, we turn off synaptic modification, reset the network, and begin testing. The test input is the 8 time step long pattern representing the CS. Then the simulation runs freely, without any external input.

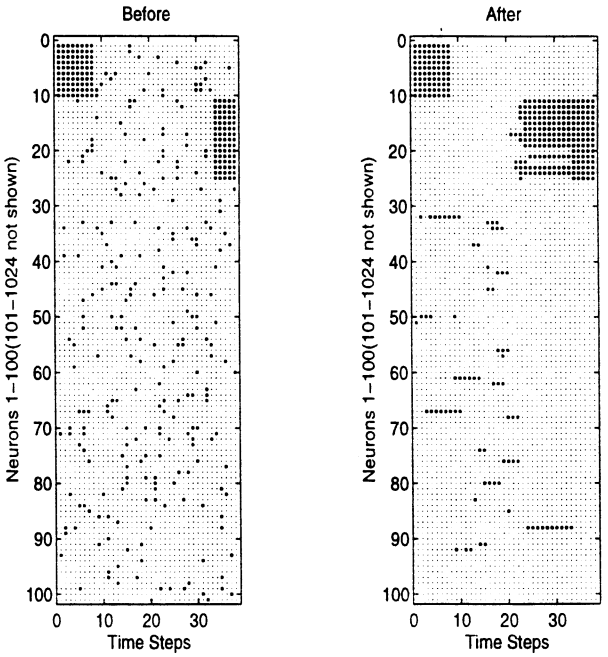

Figure 1. Sample Input. This figure contains examples of the CA3 network states for the first (left) and the last (right) trial of training. The large dots represent firing events. During the first eight time steps (160ms), neurons 1-10 are externally activiated(the CS). A 25 time step (500ms) trace period, during which there is no external stimulation, follows the CS. During the last five time steps, neurons11-25 are externally activiated (the UCS). Note the random firing of the neurons other than those directly activated by the external input at the beginning of training. After training, neurons no longer fire randomly, rather they fire for short episodes similar to place cell firing in the hippocampus. Furthermore, note there appears an anticipatory UCS with training, marked by the recurrent activation of neurons 11 to 25 around time step 22.

To examine the network code during and after training, we compare network states by calculating the cosine of the angle between the vectors representing the network states at each time step. Thus, the normalized dot product over time between two states results in a similarity matrix with dimensions of the length of one sequence by the length of the other. We use this cosine method to quantify both the

network code similarity to itself over various training cycles and to compare the network's output code to its final training code when testing. A network mediated anticipatory UCS is defined as a cell firing patter similar to and preceeding the UCS firing patterns. If properly timed relative to the duration of an eye blink, the onset of this anticipatory UCS has the information needed to produce an appropriate CR.

3. Results

Figure 1 shows the firing pattern of neurons 1 to 100 in the model for the first trial of learning and for the last trial. The training induced the alteration in firing of the recurrently activated neurons is quite apparent. The initial scattered, almost random firing, has been changed into very selective firing patterns. Specifically, a neuron remains silent for a long period of time, then it fires (if the neuron fires at all) for a short period of time, and then it remains silent. Also evident after training is the fact that the externally driven UCS neurons are activated by the recurrent connections earlier than they normally would be turned on by the external input. Indeed, we find that the activation is about 160ms earlier when the trace interval is 500ms.

Figure 2 shows the relationship, after learning, between the trace interval used during training and the early onset of the code for the unconditioned stimulus. Two observations should be made. First of all, for a 500ms trace interval in actual animal experiments, the average conditioned response is delivered 140ms prior to the onset of the UCS [5]. If we take into account some delays within the nervous system, it seems sensible that any system within the brain that is clocking the trace interval in order to produce the early onset of a conditioned response, should actually precede this 140ms value. Thus, the timing of the anticipatory UCS is certainly within a reasonable range of what might be needed by the animal. Secondly, the long trace interval paradigms are poorly, if at all, solved by the network (again Figure 2). This failure to learn the trace interval is particularly apparent for the 2000ms trace paradigm. In this case the learned response, based on the coding of the UCS, would be given almost immediately after the CS. In fact, this 2000ms is an interval that the mammalian organism cannot learn. Finally, we note that at shorter and shorter intervals, it gets more and more difficult to produce a UCS code early enough. However, to our knowledge this problem has not been investigated by experimental neuroscientists. Furthermore, we suspect that with shorter and shorter trace intervals, the role of the hippocampus in timing becomes less and less important so that eventually the cerebellum alone is able to produce the correct and appropriately timed response.

In Figure 3, we show the development of a neuronal code over trials. Early in training, there is no stable code for the trace period; i.e., the codes over time are equally dissimilar,

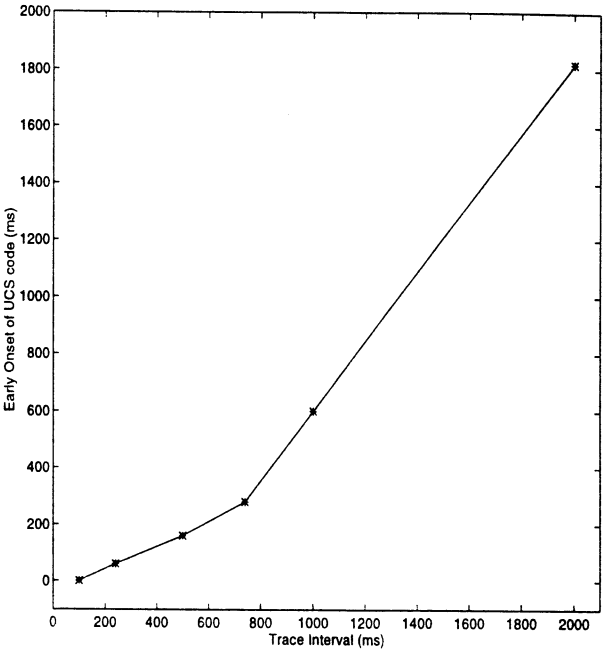

Figure 2. Early Onset of the Unconditioned Response Codes. In analogy to the psychological experiments, we tried training the network with a variety of trace intervals (indicated on the abscissa). Plotted here is the difference between the onset of the UCS and the onset of the learned cell firing patterns that code the UCS at each of the trace intervals. The UCS codes always precede the actual UCS. Note that the network trained in the 500ms trace paradigm produces the code for the UCS 160ms prior to the actual arrival of the UCS. This interval seems to fit experimental observations. Also fitting experimental observations are the much too early anticipatory UCS's when the trace interval is substantially lengthened. Networks trained with shorter trace intervals do not produce the UCS code much before the arrival of the actual UCS.

denoted by the light gray texture. Gradually, the network forms codes for the CS and the UCS. Then the network begins to form a code for the trace interval, propagating forward from the CS. A little later, and to a lesser extent, the UCS code moves backward. The forward and backward propagating codes meet around trial 200, at which point we can say that the network has fully encoded the sequence.

Simply forming a code for both the CS and the UCS is not enough for the network to recall the UCS when prompted

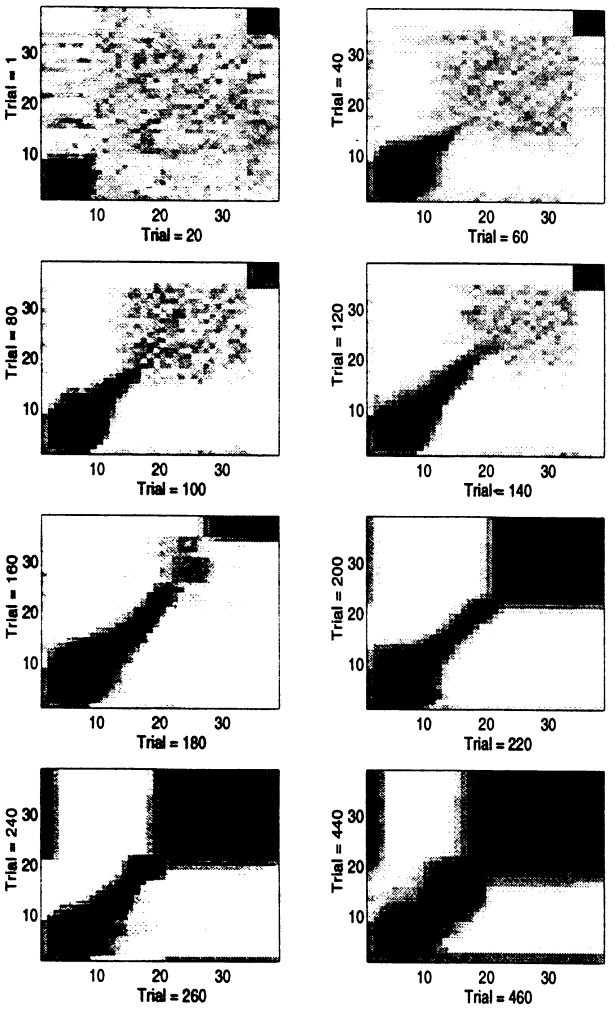

Figure 3. Code Formation. Over learning trials of CS-trace-UCS sequences, an internal code develops in CA3. Shown here are pattern similarities during the course of training. Each panel compares the 38 sequential CA3 cell firing patterns for a pair of training trials. The darker the square, the stronger the similarity. At the beginning of training, the CS and the UCS codes are only crudely distinguished from the trace interval and are changing across trials (trial 1 vs. trial 20). By trial 40, the distinct representations for the CS and UCS are formed as are a few trace bridging patterns growing out of the CS encoding. By trial 180 the code for the UCS has moved backward several time steps and a nearly complete sequence of moderately distinct patterns bridges the trace interval. By trial 200 the trace interval is bridged by a relatively stable code. The external input is supplied over time steps 1-8(the CS) and time steps 34-38(the UCS). In this simulation the trace period is 25 time steps (500ms).

with the CS. If we test the network early in training, i.e. before it has formed a complete code for the trace interval connecting the CS to the UCS, the network can not anticipate the UCS. Thus, the network must form a complete code across the trace interval to bridge the CS to the anticpatory UCS.

The network was able to form a code spanning the trace intervals of 100ms, 240ms, 500ms, 740ms, and 1000ms. With trace intervals of 500ms and above, however, the network's code for the UCS encompassed all the patterns beyond 400ms into the trace period. Therefore, there seems to be a limit to the length of the trace interval the network can span before the UCS expands too far back. When presented with the CS and UCS separated by a 2000ms trace interval, the network did not form any distrinctive code for the trace interval, coding the entire sequence with a single pattern of neurons that fire throughout the training cycle, changing only with the external input.

4. Discussion

Trace classical conditioning is an extremely simple form of learning, which is one of the reasons neuroscientists find it such an interesting paradigm to investigate. Two other reasons make this paradigm interesting. First, this paradigm is clearly nonspatial in nature, the CS's are usually auditory; therefore, the cognitive mapping theory of hippocampal function is irrelevant to understanding the role of the hippocampus in trace conditioning. Secondly, the requirement for the hippocampus in trace conditioning is only over the initial few hundred trials [5]. Thus, hippocampal function in trace conditioning is highly reminiscent of the function posited for humans. That is, the neocortex is unable to learn and store new memories in a long-term fashion without a hippocampus that acts as an intermediate storage device. For all these reasons a good neural network model of the hippocampus should reproduce and explain the role of the hippocampus in trace conditioning and eventually make experimental predictions. Here we are hypothesizing that the correct timing that warns the animal of the impending UCS is at least initially encoded by the hippocampus.

Although many variables remain to be investigated, we are very much encouraged by the fact that the code for the UCS migrates somewhat earlier than the UCS itself. We are also encouraged by the fact that long trace intervals, which an animal cannot learn [11], are also not learned by the

network. With the establishment of these results, we may have a suitable model to understand the cellular and synaptic functions that determine how far back in time the UCS code will migrate and what processes prevent the learning of longer trace intervals. Also, if we build more physiological models, we should be able to predict even more detailed cell firing patterns.

Still, and even without these more physiological models, data, such as that seen in right portion of Figure 1, predict a type of cell firing previously called local context neuronal firings [9, 6, 13]. In these patterns, neurons essentially take turns— and only one turn to a neuron— in marking a particular subinterval. As in our earlier work with sequence learning, where individual neurons learned subsequences, here the cell by cell learning is of subintervals even though there is no input pattern to learn during the trace interval. If we are correct and such firing patterns prove critical, then one limitation on trace learning is activity level and its implied effects on sequence length capacity [9].

References

[1] D. G. Amaral, N. Ishizuka, and B. Claiborne. Neurons, numbers, and the hippocampal network. In J. Storm-Mathisen, J. Zimmer, and O. P. Ottersen, editors, *Understanding the Brain Through the Hippocampus: The Hippocampal Region as a Model for Studying Structure and Function*, pages 1–11. Elsevier, 1990.

[2] B. Gustafsson and H. Wigstrom. Hippocampal long-lasting potentiation produced by pairing single volleys and brief conditioning tetani evoked in separate afferents. *J. Neurosci.*, 6(6):1575–1582, 1986.

[3] W. R. Holmes and W. B. Levy. Insights into associative long-term potentiation from computational models of NMDA receptor-mediated calcium influx and intracellular calcium concentration changes. *J. Neurophys.*, 63(5):1148–1168, 1990.

[4] N. Ishizuka, J. Weber, and D. G. Amaral. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.*, 295:580–623, 1990.

[5] J. J. Kim, R. E. Clark, and R. F. Thompson. Hippocampectomy impares the memory of recently, but not remotely, acquired trace eyeblink conditioned responses. *Behavioral Neuroscience*, 109(2):195–203, 1995.

[6] W. B. Levy. Unification of hippocampal function via computational considerations. In *INNS World Congress on Neural Networks*, volume 4, pages 661–666, New Jersey, 1994. Lawrence Erlbaum.

[7] W. B. Levy and O. Steward. Synapses as associative memory elements in the hippocampal formation. *Brain Res.*, 175:233–245, 1979.

[8] W. B. Levy and O. Steward. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–797, 1983.

[9] W. B. Levy and X. B. Wu. The relationship of local context codes to sequence length memory capacity. *Network*, 1996.

[10] J. R. Moyer, R. A. Deyo, and J. F. Disterhoft. Hippocampectomy disrupts trace eye-blink conditioning in rabbits. *Behavioral Neuroscience*, 104:243–252, 1990.

[11] P. R. Solomon, R. F. Thompson, and D. J. Weisz. Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behav. Neurosci.*, 100(5):729–744, 1986.

[12] R. F. Thompson and J. E. Steinmetz. *Learning and memory: The behavioral and biological substrates.*, chapter The essential memory trace circuit for a basic form of associative learning., pages 369–386. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, US, 1992.

[13] X. B. Wu, R. A. Baxter, and W. B. Levy. Context codes and the effect of noisy learning on a simplified hippocampal ca3 model. *Biol. Cybern.*, 74:159–165, 1995.

Acknowledgments

This work was supported by NIH MH48161 and MH00622, EPRI RP8030-08, and Pittsburgh Supercomputing Center Grant BNS950001 to WBL, and by the Department of Neurosurgery, Dr. John A. Jane, Chairman.

The 1997 IEEE International Conference on Neural Networks

June 9 - 12, 1997
Westin Galleria Hotel,
Houston, Texas, USA

The 1997 IEEE International Conference on Neural Networks Proceedings

IEEE Catalog Number: 97CH36109
ISBN Softbound: 0-7803-4122-8
ISBN Casebound: 0-7803-4123-6
ISBN Microfiche: 0-7803-4124-4
Library of Congress: 97-72856

Additional Proceedings may be ordered from:

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
(800) 678-IEEE