Presented at 11th International Conference on Neural Information Processing
(ICONIP2004), Kolkata, India, Nov. 22–25, 2004

A neural network model for trace conditioning

Tadashi Yamazaki and Shigeru Tanaka

Laboratory for Visual Neurocomputing, RIKEN Brain Science Institute,
2-1 Hirosawa, Wako, Saitama 351-0198, JAPAN
tyam@brain.riken.jp, shigeru@riken.jp

Abstract. We studied the dynamics of a neural network which have both of recurrent excitatory and random inhibitory connections. Neurons started to become active when a relatively weak transient excitatory signal was presented and the activity sustained due to the recurrent excitatory connections. The sustained activity stopped when a strong transient signal was presented or when neurons were disinhibited. The random inhibitory connections modulated the activity patterns of neurons so that the patterns evolved without recurrence with time. Hence, a time passage between the onsets of the two transient signals was represented by the sequence of activity patterns.
We then applied this model to the trace eyeblink conditioning which is mediated by the hippocampus. We assumed this model as CA3 of hippocampus and considered an output neuron corresponding a neuron in CA1. The activity pattern of the output neuron was similar to that of CA1 neurons during the trace conditioning which was experimentally observed.

1 Introduction

It is widely known that the hippocampus plays a critical role in declarative learning and memory. An example of declarative learning is trace eyeblink conditioning [1]. In this paradigm, a conditioned stimulus (CS; e.g., a tone) presented before an unconditioned stimulus (US; e.g., an airpuff) which elicits an automatic conditioned response (CR; e.g., an eyeblink). The offset of the CS and onset of the US do not overlap, creating a off-stimulus interval. The subject is tested for learning an association between the CS and the US, as evidenced by the CR in anticipation of the US. The hippocampus is known to play a necessary role in learning a well-timed CR in anticipation of the US [2]. Thus, neurons in the hippocampus must have a kind of "memory trace" of the CS that bridges the time interval to form a CS-US association, but how?

So far, we have studied a simple random recurrent inhibitory network and found that activity patterns of neurons evolved with time without recurrence due to random recurrent connections among neurons. The sequence of activity patterns was generated by the trigger of an external signal, suggesting that a time passage from the trigger of an external signal could be represented by the sequence of activity patterns [3]. In this paper, we extended this idea and

studied the dynamics of a neural network model which have both recurrent excitatory and random inhibitory connections. When a relatively weak transient signal was presented, neurons in this model started to become active and the activity was sustained even during the off-stimulus period due to the recurrent excitatory connections. The sustained activity stopped when a strong transient signal was presented or when neurons were disinhibited. On the other hand, since the random recurrent connections generated non-recurrent activity patterns of neurons, the time passage was represented.

We then applied this model to the trace eyeblink conditioning which is mediated by the hippocampus. We assumed this model as CA3 of hippocampus and considered an output neuron corresponding a neuron in CA1. The activity pattern of the output neuron was similar to that of CA1 neurons during the trace conditioning which is experimentally observed.

2 Model description

The model consists of N excitatory neurons and the same number of inhibitory neurons. Let $z_{\mathrm{ex}i}(t)$ and $z_{\mathrm{inh}i}(t)$ be the activities of excitatory neuron i and inhibitory neuron i at time t, respectively. For a neuron type $\mathrm{T} \in \{\mathrm{ex}, \mathrm{inh}\}$, $z_{\mathrm{T}i}(t)$ is defined as

$$z_{\mathrm{T}i}(t) = \begin{cases} u_{\mathrm{T}i}(t) & u_{\mathrm{T}i}(t) > \theta_{\mathrm{T}}, \\ 0 & \text{otherwise.} \end{cases}$$

$u_{\mathrm{ex}i}(t)$ and $u_{\mathrm{inh}i}(t)$ are internal states of excitatory neuron i and inhibitory neuron i at time t, respectively, which are calculated as

$$\tau_{\mathrm{ex}} \dot{u}_{\mathrm{ex}i}(t) = -u_{\mathrm{ex}i}(t) + I_i(t) + \sum_j w_{\mathrm{ex}i \leftarrow \mathrm{ex}j} z_{\mathrm{ex}j}(t) - \sum_j w_{\mathrm{ex}i \leftarrow \mathrm{inh}j} z_{\mathrm{inh}j}(t) \quad (1)$$

$$\tau_{\mathrm{inh}} \dot{u}_{\mathrm{inh}i}(t) = -u_{\mathrm{inh}i}(t) + \sum_j w_{\mathrm{inh}i \leftarrow \mathrm{ex}j} z_{\mathrm{ex}j}(t) - \sum_j w_{\mathrm{inh}i \leftarrow \mathrm{inh}j} z_{\mathrm{inh}j}(t), \quad (2)$$

where for $\mathrm{T}, \mathrm{T}' \in \{\mathrm{ex}, \mathrm{inh}\}$, $w_{\mathrm{T}i \leftarrow \mathrm{T}'j}$ is the weight of the synaptic connection from neuron j of type T' to neuron i of type T, τ_{T} is the time constant, and $I_i(t)$ is the external input to excitatory neuron i at time t.

Synaptic connections are defined as follows. $w_{\mathrm{ex}i \leftarrow \mathrm{ex}j}$ is set at $c_{\mathrm{ex} \leftarrow \mathrm{ex}}$ for any i and j: excitatory connections are all-to-all, $w_{\mathrm{inh}i \leftarrow \mathrm{ex}j}$ is given under the binomial distribution $\Pr(w_{\mathrm{inh}i \leftarrow \mathrm{ex}j} = 0) = \Pr(w_{\mathrm{inh}i \leftarrow \mathrm{ex}j} = c_{\mathrm{inh} \leftarrow \mathrm{ex}}/N) = 0.5$, $w_{\mathrm{ex}i \leftarrow \mathrm{inh}j}$ is set at $c_{\mathrm{ex} \leftarrow \mathrm{inh}}$ if $i = j$ and 0 otherwise, thus each inhibitory neuron inhibits its corresponding excitatory neuron, and $w_{\mathrm{inh}i \leftarrow \mathrm{inh}j}$ is set at $c_{\mathrm{inh} \leftarrow \mathrm{inh}}$ for any i and j for simplicity. When we consider disinhibition of excitatory neurons, $c_{\mathrm{ex} \leftarrow \mathrm{inh}}$ and $c_{\mathrm{inh} \leftarrow \mathrm{inh}}$ are set at the half.

External input signals are given as follows. For any t and i, $I_i(t)$ is set at I_{aff} when $1 \leq i \leq N_{\mathrm{CS}}$ and $t_{\mathrm{CSonset}} \leq t \leq t_{\mathrm{CSoffset}}$, I_{aff} when $1 \leq i \leq N_{\mathrm{US}}$ and $t_{\mathrm{USonset}} \leq t \leq t_{\mathrm{USoffset}}$, and 0 otherwise.

We demonstrate that the activity patterns of neurons generated using Eq. (2) can represent a time passage, that is, the activity pattern at one time step is

dissimilar to the pattern at a different time step when the interval between the two steps are large. Therefore, we use the following correlation function as the similarity index.

$$C(t_1, t_2) = \frac{\sum_i z_{\mathrm{exi}}(t_1) z_{\mathrm{exi}}(t_2)}{\sqrt{\sum_i z^2_{\mathrm{exi}}(t_1)} \sqrt{\sum_i z^2_{\mathrm{exi}}(t_2)}}. \tag{3}$$

Parameter values are set arbitrarily as follows: $N = 1000$, $T = 1000$, $\tau_{\mathrm{ex}} = 20.0$, $\tau_{\mathrm{inh}} = 50.0$, $\theta_{\mathrm{ex}} = 0.1$, $\theta_{\mathrm{inh}} = 0.1$, $c_{\mathrm{ex}\leftarrow\mathrm{ex}} = 3.0$, $c_{\mathrm{inh}\leftarrow\mathrm{ex}} = 6.0$, $c_{\mathrm{ex}\leftarrow\mathrm{inh}} = 20.0$, $c_{\mathrm{inh}\leftarrow\mathrm{inh}} = 6.0$, $I_{\mathrm{aff}} = 1.0$, $N_{\mathrm{CS}} = 400$, $N_{\mathrm{US}} = 1000$, $t_{\mathrm{CSonset}} = 0$, $t_{\mathrm{CSoffset}} = 20$, $t_{\mathrm{USonset}} = 800$, and $t_{\mathrm{USoffset}} = 820$.

3 Results

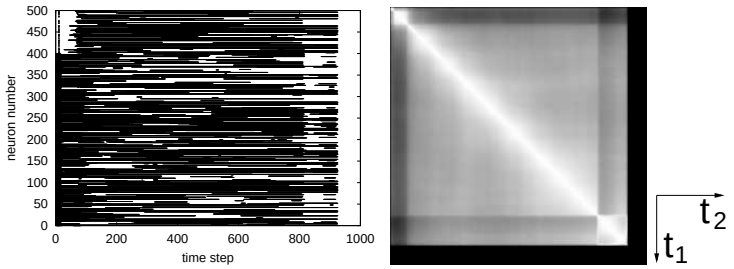

Fig. 1. Raster plot of active states ($z_{\mathrm{exi}}(t) > \theta_{\mathrm{ex}}$) of the first 500 excitatory neurons (left) and the similarity index (right).

The left panel in Fig. 1 shows the plot of active states of the first 500 neurons out of N excitatory neurons during T steps. At time t ($t_{\mathrm{CSonset}} \leq t \leq t_{\mathrm{CSoffset}}$), neurons i ($1 \leq i \leq N_{\mathrm{CS}}$) were given external signals I_{aff} and they started to become active. Then, the activity spread out among all neurons through recurrent excitatory connections and was sustained during the off-stimulus period ($t > t_{\mathrm{CSoffset}}$). At time t ($t_{\mathrm{USonset}} \leq t \leq t_{\mathrm{USoffset}}$), neurons i ($1 \leq i \leq N_{\mathrm{US}}$) were given external signals but due to the recurrent inhibition not all neurons became active. After the USoffset, neurons gradually became inactive and suddenly stopped activities at $t \approx 920$. We also examined the effect of disinhibition. This time, we did not present the external signals but disinhibited neurons during ($t_{\mathrm{USonset}} \leq t \leq t_{\mathrm{USoffset}}$). After the disinhibition, the sustained activity suddenly stopped (data not shown). During the off-stimulus period, once a neuron started to become active, its activity continued for several hundreds of steps, and then the neuron became inactive. Some neurons were reactivated after the inactive period. Thus, the active and inactive periods appeared alternately.

The right panel in Fig. 1 shows the similarity index calculated using Eq. (3). Since Eq. (3) takes two arguments of t_1 and t_2, we obtained a $T \times T$ matrix, where the row and the column were specified by t_1 and t_2, respectively. Similarity indices were plotted in a gray scale in which black indicated 0 and white 1. A white band appeared diagonally. Since the similarity index at the identical step ($t_2 = t_1$) takes 1, the diagonal elements of the similarity index appeared white. The similarity index decreased monotonically as the interval between t_1 and t_2 became longer. This result indicates that the activity pattern of neurons changed gradually with time and did not recur.

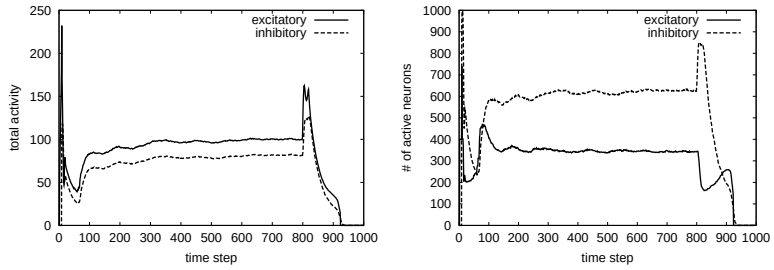

Fig. 2. Total activity of excitatory (solid line) and inhibitory (dashed line) neurons (left) and the number of active excitatory (solid line) and inhibitory (dashed line) neurons (right).

The left panel in Fig. 2 shows the total activity of neurons while the right panel the number of active neurons. As can be seen, both of the total activity and the number of active neurons remained constant during the off-stimulus period. These results suggest that the representation of a time passage was stable.

Then, we examined if this model could make association between the CS and US onsets temporally separated, which were motivated by the trace eyeblink conditioning. We regarded the present model as the hippocampus CA3 because of its recurrent excitatory and inhibitory connections and we considered an output neuron corresponding a neuron in CA1, which was connected with excitatory neurons in the model with synaptic weights representing Schaffer collaterals. We assumed that the output neuron received only the US directly through an another pathway corresponding to the perforant path and the CA3 neurons received only the CS (see Discussion). The output neuron learns to associate the US onset with the CS onset which are disconnected by the off-stimulus interval [4].

We ran the simulation twice. In the first run, we determined active neurons at $t = t_{\mathrm{USonset}}$ and set their synaptic weights to 1 while weights of inactive neurons 0, namely,

$$w_i = \begin{cases} 1 & z_{\mathrm{ex}i}(t_{\mathrm{USonset}}) > \theta_{\mathrm{ex}}, \\ 0 & \text{otherwise}, \end{cases}$$

where w_i represents the synaptic weight of neuron i. This corresponds to the Long-Term Potentiation (LTP) at Shaffer collaterals induced by the conjunctive stimulation of the output neuron by the US and by the signals from our CA3 neurons. In the second run, we calculated the net input to the output neuron as $\sum_i w_i z_{\mathrm{exi}}(t)$. In order to see if the output neuron anticipates the US onset only by the CS stimulation, the US signal was not presented to the output neuron. We regarded the value of the net input as the activity of the output neuron by assuming the linear response of the output neuron. Thus, the output neuron learns to elicit responses at $t = t_{\mathrm{USonset}}$ by the trigger of the CS onset $t = t_{\mathrm{CSonset}}$ followed by the off-stimulus period.

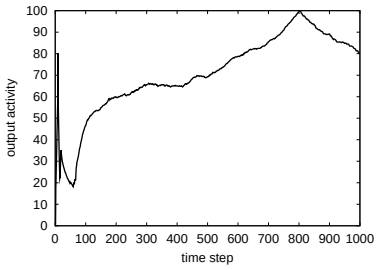

Fig. 3. Plot of the net input $\sum_i w_i z_{\mathrm{exi}}(t)$.

Figure 3 shows the activity of the output neuron. As can be seen, after the CS onset, the activity transiently increased. It reached the maximal value at $t = 10$ and then sharply decreased. After the CS offset the activity increased and converged to constant at $t \approx 300$. Then, at $t \approx 500$ it started to move again and slowly increased towards $t = t_{\mathrm{USonset}}$. At $t = t_{\mathrm{USonset}}$ the activity reached the maximum value and then gradually decreased. As as result, the output neuron could associate the US onset with the CS onset. The initial transient increase and the following slow increase towards $t = t_{\mathrm{USonset}}$ are typical responses of neurons in the hippocampus CA1 during the trace conditioning [2]. Hence, the present model successfully reproduced the hippocampal activity in the trace conditioning.

4 Discussion

We studied the dynamics of a model which have recurrent excitatory and random inhibitory connections. Due to the recurrent excitatory connections individual neurons could generate sustained activity during the off-stimulus period while due to the random recurrent inhibitory connections the neurons exhibited the random repetition of transition between active and inactive states. Hence, the population of active neurons changed gradually with time and did not recur. This property was confirmed by calculating the similarity index.

We then examined if the present model could account for the trace eyeblink conditioning mediated by the hippocampus. We regarded the model as the hippocampus CA3 and incorporated an output neuron corresponding to a neuron in CA1. We calculated the activity of the output neurons and the activity profile was similar to the one observed experimentally [2]. We assumed that the output neuron received only the US directly through an another pathway corresponding to the perforant path. Input signals to the hippocampus first arrive at both the layers 2 and 3 of the entorhinal cortex (EC2, EC3), and neurons in CA1 receive input signals from EC3 through the perforant path and from CA3 through Shaffer collaterals [5]. Activation of perforant path neurons to CA1 do not evoke neuronal activity in CA1 when neurons in CA3 are not activated [6]. Since neurons in CA3 are inactive just before the CS onset, CA1 neurons cannot become active by the CS stimulation only. Hence, we could ignore the CS presented to the output neuron. We also assumed that the CA3 neurons received only the CS. CA1 neurons excite inhibitory neurons in the septum and in turn these inhibitory neurons inhibit inhibitory neurons in CA3 [7]. We hypothesized that this disinhibition stops the sustained activity. Therefore, we ignored the US stimulation to CA3 neurons.

In the present model, the recurrent excitatory connections were assumed to be all-to-all. Rolls has argued that CA3 plays a role of an autoassociation memory [5]. If so, the connections should be symmetric and thus the activity pattern of neurons converges to a steady state. Since all-to-all connections is a variation of symmetric connections, hence, our assumption may not be too simplistic.

Levy and his colleagues have developed a model of the hippocampus CA3 and recently in [8] they have reported that their model successfully reproduced the activity pattern of CA1 neurons in the trace conditioning. Their model connects the CS onset to the US onset by the temporally asymmetric Hebbian learning between pairs of CA3 neurons [9]. In their model, inhibitory neurons are incorporated only to regulate the total activity of CA3 neurons. On the other hand, this study demonstrated that inhibitory neurons could work more: they modulate the activity of excitatory neurons and generate a sequence of activity patterns without recurrence, which can represent a time passage from the CS onset. This study may shed light on roles of inhibitory neurons.

References

1. Christian, K.M., Thompson, R.F. Learn. and Mem. **11** (2003) 427–455
2. McEchron, M.D., Disterhoft, J.F. J. Neurophys. **78** (1997) 1030–1044
3. Yamazaki, T., Tanaka, S. In: Society for Neuroscience Abstract. (2003)
4. McNaughton, B.L. Brain Res. Rev. **16** (1991) 202–204
5. Rolls, E.T. Hippocampus **6** (1996) 601–620
6. Bartesaghi, R., Gessi, T. Hippocampus **13** (2003) 235–249
7. Tóth, K., Borhegyi, Z., Freund, T.F. J. Neurosci. **13** (1993) 3712–3724
8. Rodriguez, P., Levy, W.B. Behav. Neurosci. **115** (2001) 1224–1238
9. Devanne, D., Gähwiler, B.H., Thompson, S.M. J. Physiol. **501** (1998) 237–247