BMC Neuroscience

Poster presentation

Open Access

A network model that can learn reward timing using reinforced expression of synaptic plasticity

Jeffrey P Gavornik*[1,2], Yonatan Loewenstein[3] and Harel Z Shouval[1]

Address: [1]Department of Neurobiology and Anatomy the University of Texas Medical School in Houston, TX, USA, [2]Department of Electrical and Computer Engineering the University of Texas. Austin, TX, USA and [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Email: Jeffrey P Gavornik* - gavornik@mail.utexas.edu

* Corresponding author

from Sixteenth Annual Computational Neuroscience Meeting: CNS*2007
Toronto, Canada. 7–12 July 2007

Published: 6 July 2007

BMC Neuroscience 2007, **8**(Suppl 2):P103 doi:10.1186/1471-2202-8-S2-P103

Recent experimental results indicate that cells within the primary visual cortex can learn to predict the time of rewards associated with visual cues [1]. In this work, different visual cues were paired with rewards at specific temporal offsets. Before training, neurons in visual cortex were active only during the duration of the visual cue. After sufficient training neurons developed persistent activity for a time period correlated with the timing of reward.

Recurrent connections in a neural network can be constructed to set a desired network time constant that is different from the time constants of the constituent neurons. However, it is not known how such a network can learn the appropriate recurrent weights. A plasticity model that is able to accomplish this must be sensitive to the timing of reward events that, at least initially, occur seconds after the activity in the network returns to its basal level. In order to learn the appropriate dynamics, this network needs to solve a temporal credit assignment problem. In our model plasticity is an ongoing process changing the recurrent synaptic weights as a function of their activity; in the absence of a reward signal this plasticity rapidly decays. External reward signals allow permanent expression of preceding plasticity events, reinforcing only those which predict the reward. As a result, the network dynamics are altered and it develops time constants correlated with the timing of different rewards. As in other reinforcement learning models the reward signal is inhibited by the network activity to produce a stable activity pattern.

We have implemented these ideas in both abstract passive integrator networks and in more realistic integrate and fire networks and obtained results that are qualitatively similar to the experimental results. Further, we examine the implications of different possible biophysical mechanisms and propose experiments to test which specific mechanism are involved.

Support: NSF CRCNS grant number 0515285.

References

1. Schuler MG, Bear MF: **Reward timing in the primary visual cortex.** *Science* 2006, **311**:1606-1609.