A COMPUTATIONAL MODEL OF REINFORCEMENT LEARNING AT BASAL GANGLIA

D. Hasan, Ö. Karabacak, N. S. Şengör
Istanbul Technical University, Faculty of Electrical Electronics Eng. Istanbul Turkey
daghanhasan@hotmail.com, ozkan2917@yahoo.com, neslihan@ehb.itu.edu.tr

AIM

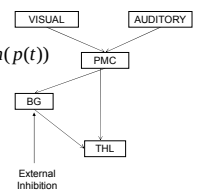
To regenerate the responses of action network
To model 'stimulus-response' association learning
To use reinforcement learning during this process

MOTIVATION

Taylor & Taylor [1,2] Action Network

- A dynamical system for BG-TH-C loop
- Temporal sequence storage and generation

$$b(t+1) = w^5 f(p(t)) + w^6 g(b(t))$$
$$p(t+1) = w^1 f(v(t)) + w^2 f(a(t)) + w^3 f(m(t)) + w^4 h(p(t))$$
$$m(t+1) = w^7 f(p(t)) + w^8 g(b(t))$$



Doya [3]

- Continuous time and space form of reinforcement learning
- Continuous Actor-Critic
- Radial Basis Neural Network
  - Value function
  - Policy function
- Minimize temporal difference(TD) errors
  expectation=reward

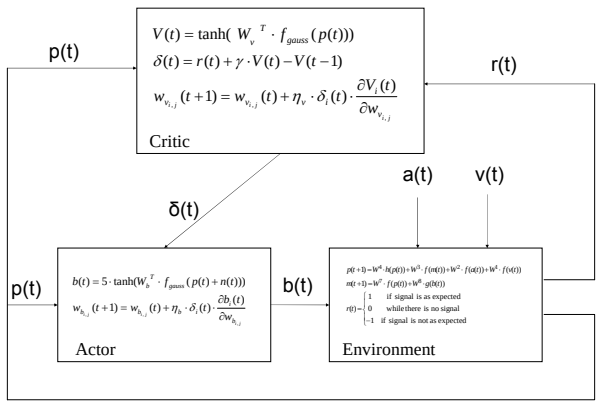
REFERENCES

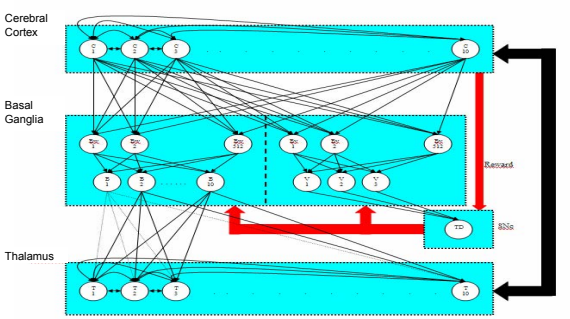
[1] Taylor, J. G., Taylor,N. R., "Analysis of recurrent cortico-basal ganglia-thalamic loops for working memory", Biological Cybernetics, Vol. 82, pp. 415-432,2000.

[2] Taylor, N.R., Taylor, J.G., "Hard-wired models of working memory and temporal sequence storage and generation", Neural Networks, Vol. 13, pp. 201-224, 2001.

[3] Doya, K., "What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?", Neural Networks, Vol. 12, pp. 961-974, 1999.

PROPOSED MODEL

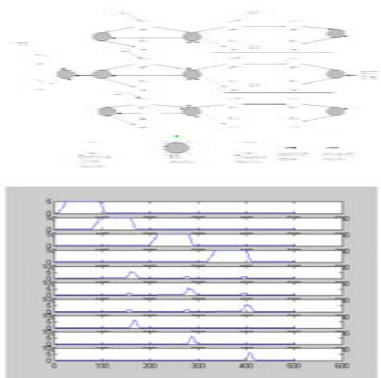




- How ~~do~~ Basal ~~Ga~~nglia correspond to actor-critic model?
  SNc δ(t) → dopamine
  Striosome → V(t)
  Matrix b(t)
- Reward is a three dimensional ~~-~~vector
  each component of the vector a motor action
  these three duplicate six more

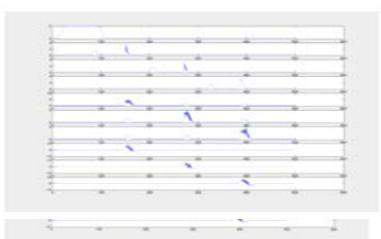
SIMULATION RESULTS

ACTION NETWORK



TRIAL 1

(r(t) is applied one time step after GO signal for ten time steps.)



TRIAL 2

(r(t) is applied simultaneously with GO signal for ten time steps.)

CONCLUSION

- Even though we could not succeed to obtain signals as action network generated, we managed to apply reinforcement signal to get similar signals.
- The signals related with movement and premovement neurons are almost same, but the signals related with simple memory are not as expected.
- If we had known more about the experiment, we could have better managed to apply reinforcement signal.
- Instead of using gardient based method it could have been better to use Hebb rule.