Real Time Learning of Rearranging Associations on a Humanoid Robot

Emeç Erçelik, Neslihan Serap Şengör

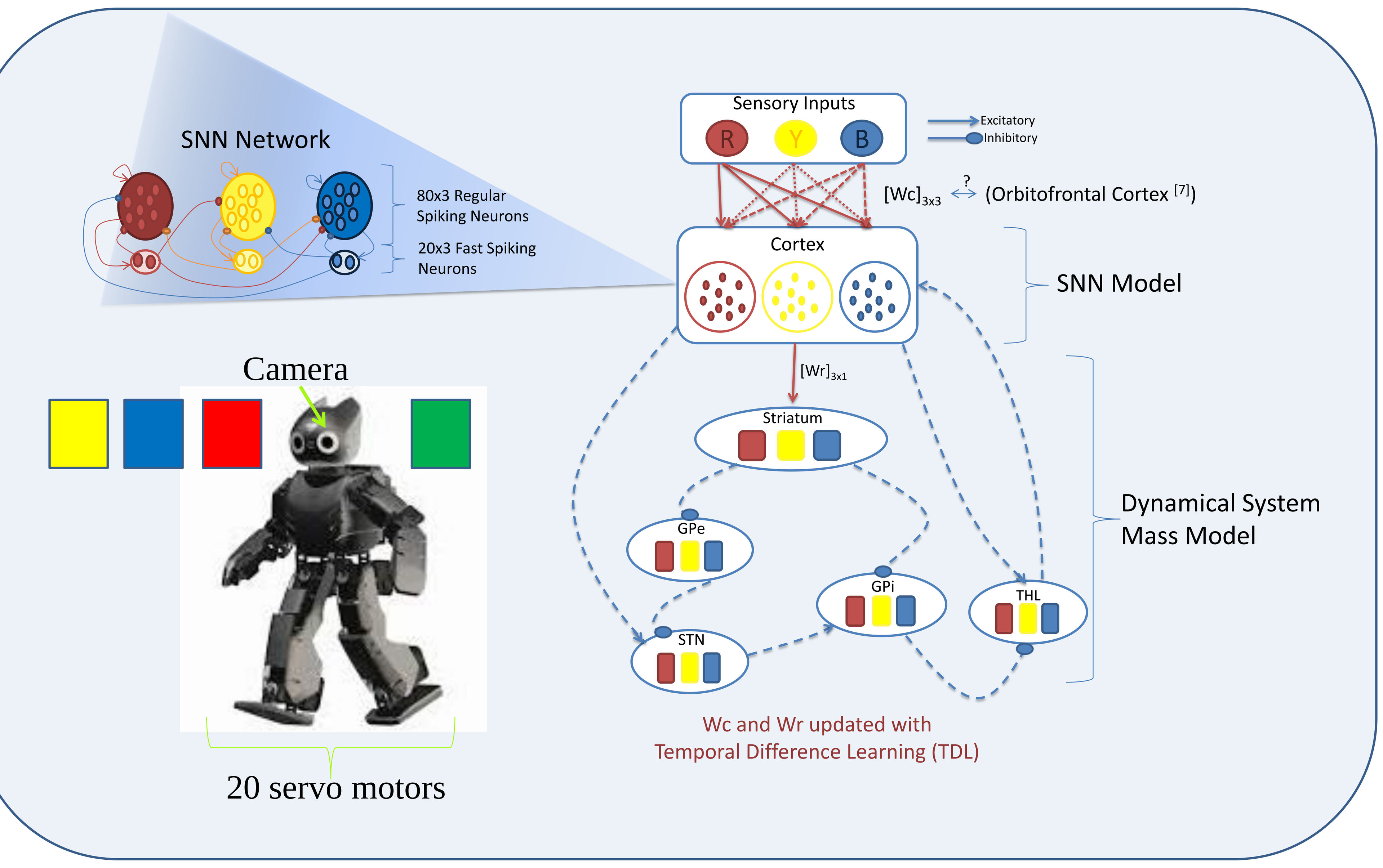
Electronics and Communication Engineering Department, İstanbul Technical University, {ercelike,sengorn}@itu.edu.tr

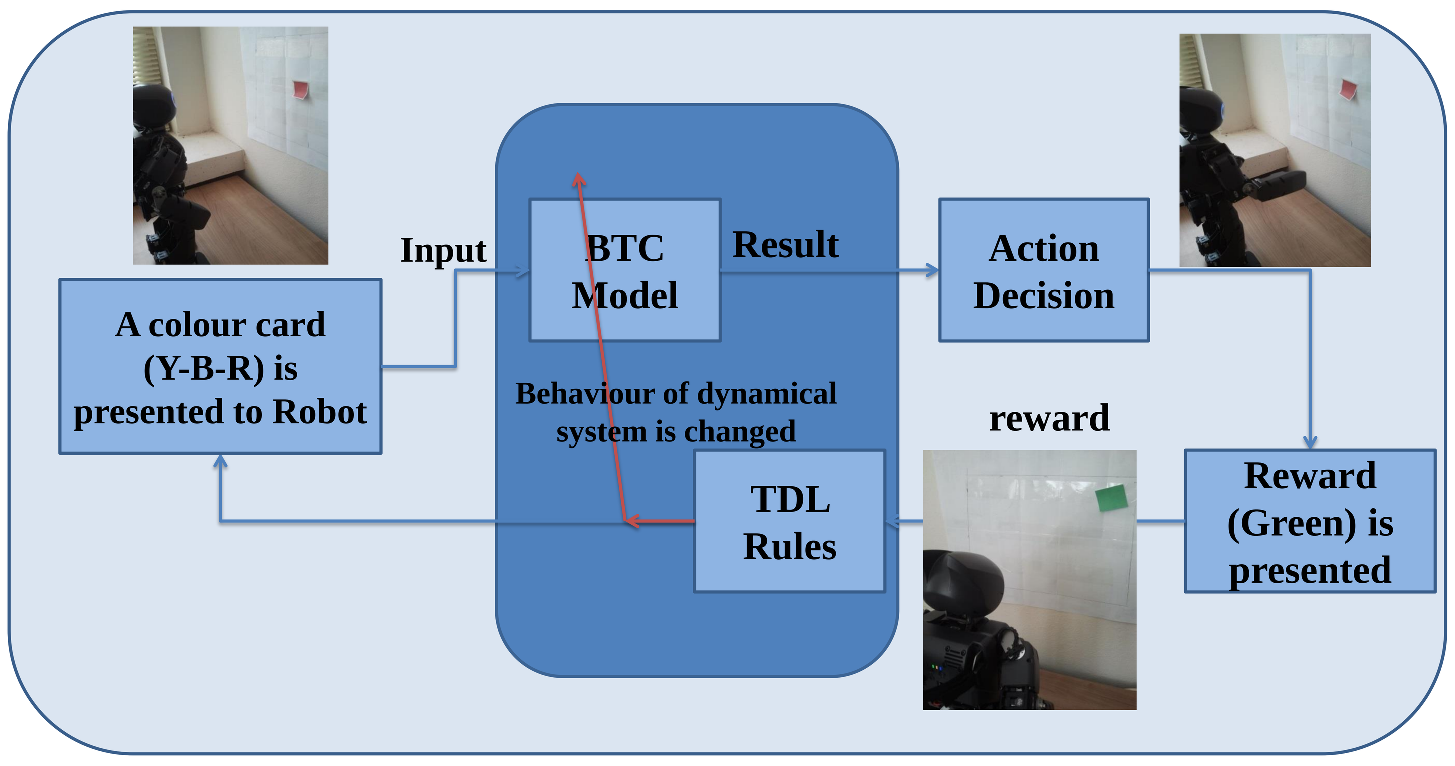
www.simmag.itu.edu.tr

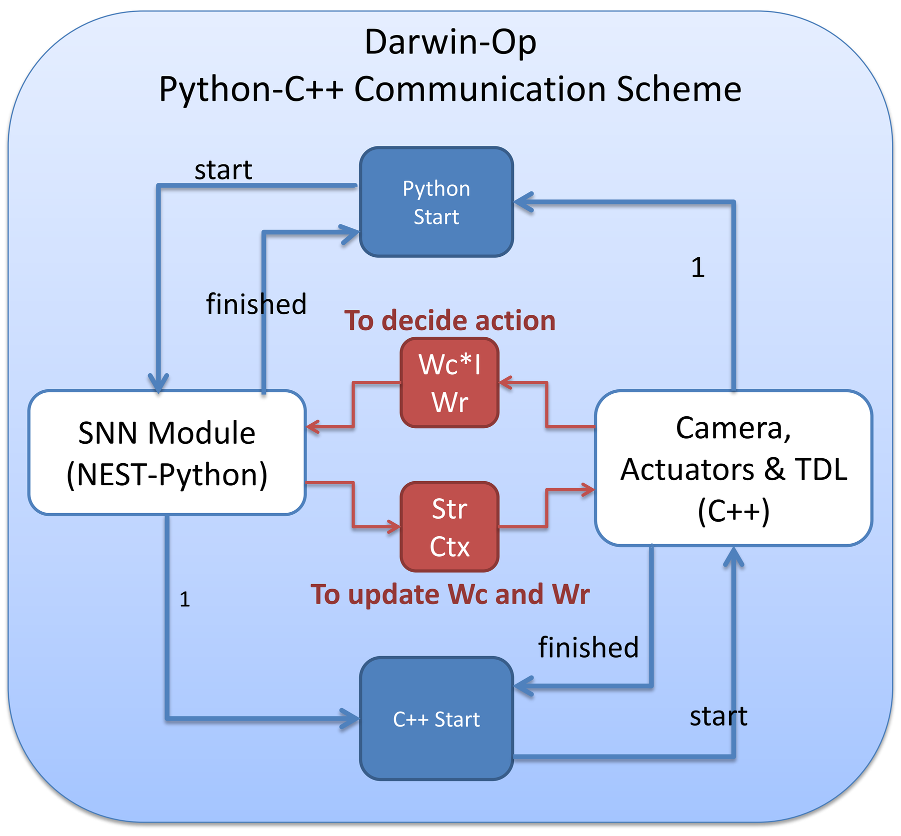
Bernstein Conference 2014, Göttingen

Abstract

In this study, we focused on modeling a temporal sequence task performed by macaque monkeys on a humonoid robot. In the task, it expected to **match** a stimulus with an appropriate action [1] and to **rearrange** the association built up. We utilized a computational model of **Cortex-Basal ganglia-Thalamus** loop to select appropriate action and **reinforcement learning** approach to realize updating model parameters **(Wc & Wr)** with temporal difference method[2,3]. Humanoid robot, Darwin-Op is used for implementation. The computational model is constructed using **Spiking Neural Network** for the Cortex part and using **dynamical system mass model** for Basal Ganglia-Thalamus part. This study is aimed to be a step to close the gap between the neuron based approaches and behavioral approaches used to realize learning in robotic applications. The task is structured upon matching a color with a desired action and rearranging the built up association to select a new action using reward.







The Model

The model consists of two parts, one is Cortex that is modeled by using SNN[4] and the other is Basal ganglia-Thlamus part that is modeled by using dynamical system approach [5]. The BTC model is coded in Python and the Cortex is simulated in NEST[6]. There are three channels each for a different action. And the winning channel in Cortex is considered as the selected action.

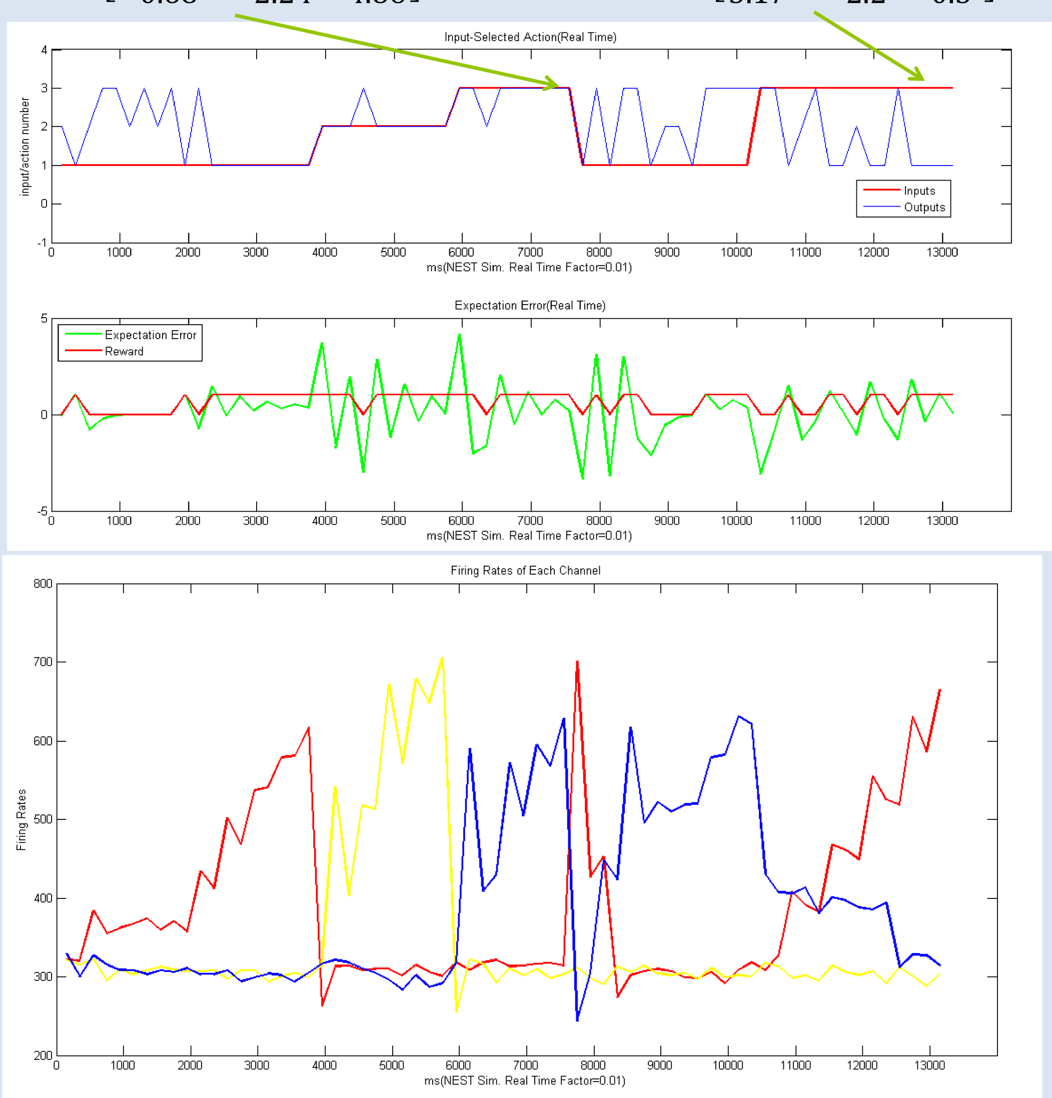
The camera information is get from the C++ coded part and sent to the model to get action decision. The output of Cortex is modeled by using firing rates and this information sent to the Basal ganglia-Thalamus loop in a normalized form to manage the calculations. At the end of this calculation the output of Ctx is sent with Striatum information of model (to be use in RL) to update the model parameters with using TDL in C++ coded part.

The Task

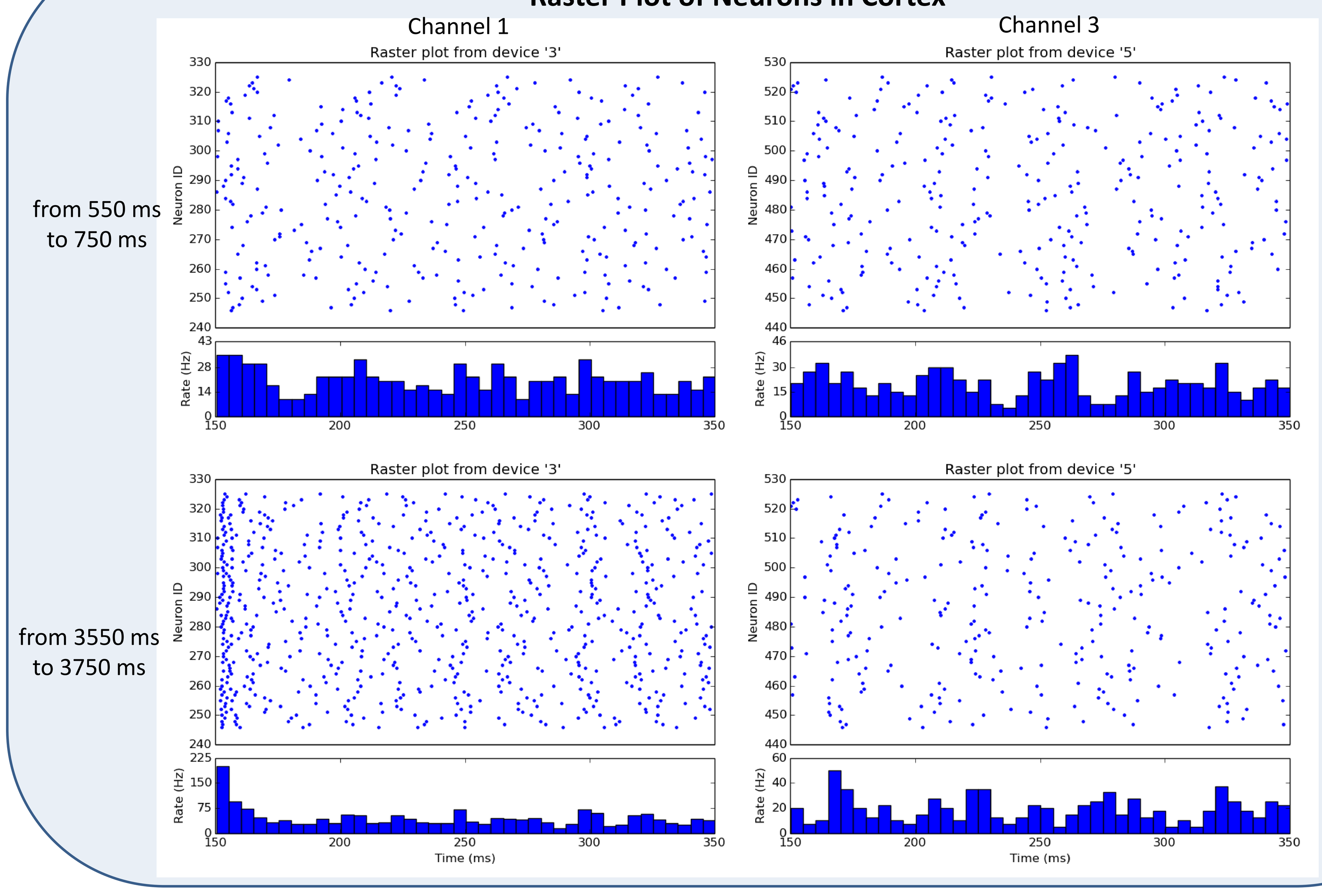
As shown in the left diagram, when a color is presented to robot, it tries to decide on an action, at first randomly, and gets reward or not according to its decision. The reward given, while the robot was **expecting nothing** as a result of its choice, causes an **expectation error**. This urges robot to **reinforce the desired stimulus-action pair**. Once robot is rewarded due to a right choice, expectation error **updates the parameters** of BTC model in charge of action selection. When update is completed, robot manages to select the desired action to a given stimulus. In case, when **an expected reward is not given** to the selected action associated with a stimulus, this also causes an expectation error and urges robot to **rearrange** the previously built association between a stimulus and action.

Real Time Results

$$Wc_{first}=\begin{bmatrix}5.3 & 0.13 & 0.17\\ -0.55 & 6.47 & -1.2\\ -0.68 & -2.24 & 4.86\end{bmatrix}\qquad Wc_{second}=\begin{bmatrix}-1.7 & 0.13 & 5.71\\ -1.1 & 6.47 & -3.1\\ 5.17 & -2.2 & 0.3\end{bmatrix}$$



Raster Plot of Neurons in Cortex



for video

http://web.itu.edu.tr/ercelike/research.html

Results

In this study, the association between different colours and actions is realized in real time on the humanoid robot platform, Darwin-Op, by using computational models and learning. So, robot manages to rearrange its associations by changing parameters of BTC model, which directly effects action selection (firing rates of Cortex) and is biologically meaningful. In robotic applications, it is seen that the rearrangement of a pair gets harder if the pair is reinforced more. As a future work, it is aimed to investigate the biological correlation between the parameters of BTC model and role of dopamine on action selection by using a more complex model of BTC loop.

Acknowledgements

* Thanks to Berat Denizdurduran for his valuable discussions.
* This work has been supported by Tübitak Project 111E264.

References

[1] Tanji J. & Shima, K., 1994. Role For Supplementary Motor Area Cells in Planning Several Movements Ahead. Nature 371, 413-416.
[2] Schultz, W., Dayan, P. and Montague, P. (1997). A Neural Substrate of Prediction and Reward, Science,275, pp.1593-1599.
[3] Denizdurduran B., Sengor N.S., 2012. A Realization of Goal-Directed Behavior-Implementing a Robot Model Based on Cortico-Striato-Thalamic Circuits. ICAART 2012, Algarve, Portugal.
[4] Izhikevich, E. M., 2003. Simple Model of Spiking Neurons, IEEE Transactions on Neural Networks, Vol. 14, No. 6, November 2003.
[5] N.S.Şengör, Ö.Karabacak, U. Steinmetz, " A Computational Model of Cortico-Striato-Thalamic Circuits in Goal-Directed Behaviour", LNCS 5163, Proceedings of ICANN 2008 328-337, 2008.
[6] Gewaltig, M.O. & Diesmann, M., 2007. NEST (Neural Simulation Tool), *Scholarpedia* 2(4):1430.
[7] Bechara, A. et al., 1997, Deciding Advantageously Before Knowing the Advantageous Strategy, Science, 275