Implementing the Learning of Appropriate Action Selection in Real Time

Emeç Erçelik*, Berat Denizdurduran**, Neslihan Serap Şengör*
*Electronics and Communication Department, Istanbul Technical University, {ercelike,sengorn}@itu.edu.tr
**Ecole Polytechnique Federale de Lausanne, Human Brain Project, {berat.denizdurduran}@epfl.ch
www.simmag.itu.edu.tr

Bernstein Sparks Workshop 2013

ABSTRACT

Cognitive robotics aims to merge the modelling studies in neuroscience with robotics to find solutions to real-time, real world challenging problems. So, inspired by models of basal ganglia circuits proposed for action selection and decision making, an implementation of such a model on a mobile robot is realized. A computational basal ganglia-thalamus-cortex (BTC) model proposed for decision making task is utilized on a mobile robot to realize foraging task. We focused on implementing the model to a robot which has a limited sensory ability to show the applicability of a simple biologically plausible model in real world problem. The environmental data, which has been taken by robot's sensors, are processed by BTC and using reinforcement learning rules embedded with BTC model, robot is urged to learn to make correct choices in a changing environment.

Cortex (Ctx), striatum (Str), globus pallidus external (GPe), subthalamic nucleus (Stn), globus pallidus internal (GPi) and thalamus (Thl) form the BTC circuits and since the work of Alexander in 1990 the role of these circuits in cognitive processes are investigated throughly in neuroscience. The connection between these structures in forming the circuit is given in Figure 1 and the Equations defining the dynamical process is given in Equation 1. The decions on actions emerge as the solution of a set of difference equations and the action selected is determined as the response of Ctx component. The evaluation of the selected action is realized according to the reinforcement rules given in Equation 2, where V and δ defines the value function andexpectation error, respectively. With updating the value function and association build between the stimulus and action selected, learning process is completed.

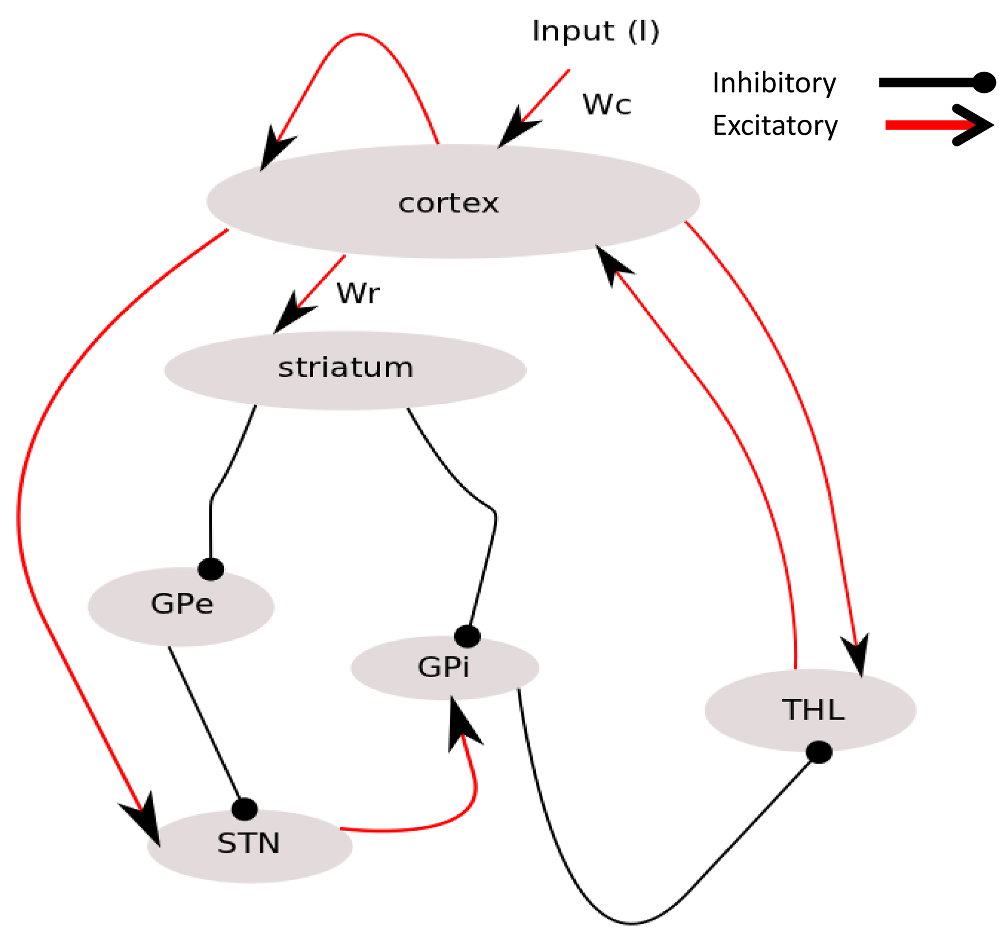


Figure 1: BTC Circuit

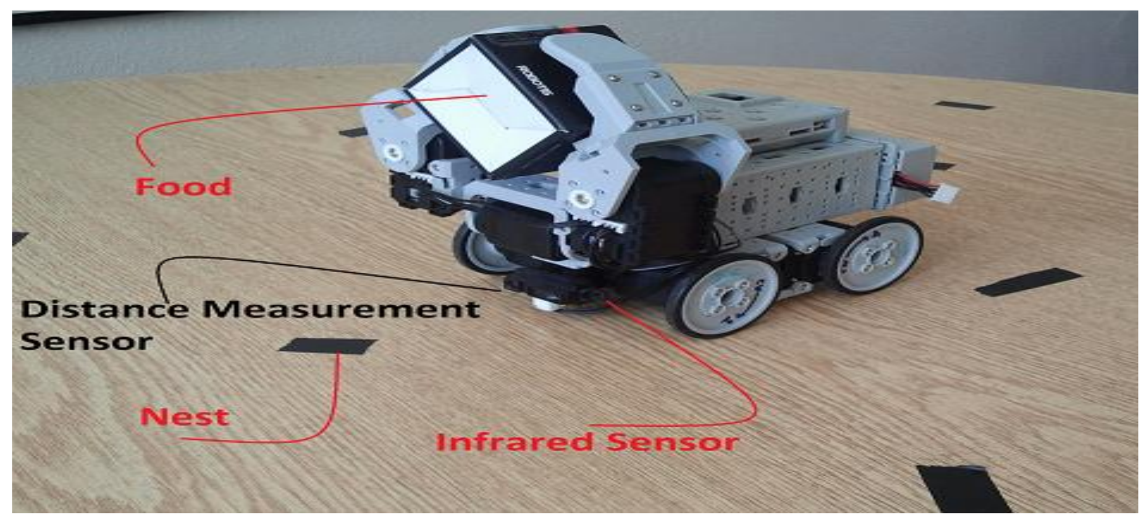


Figure 2: Mobile robot and its environment

Robot observes the external world with its Distance Measurement Sensor (DMS) and its Infrared Sensor (IR). Robot uses DMS to detect an object at its front side and to calculate the object's size. IR detects possible nest which is seen as a black band at the picture. The sensor values are first normalized to [0,1] range and then it is scaled by Pcycle, Pgrip, Pnest functions seen below.

Equations of BTC Circuit

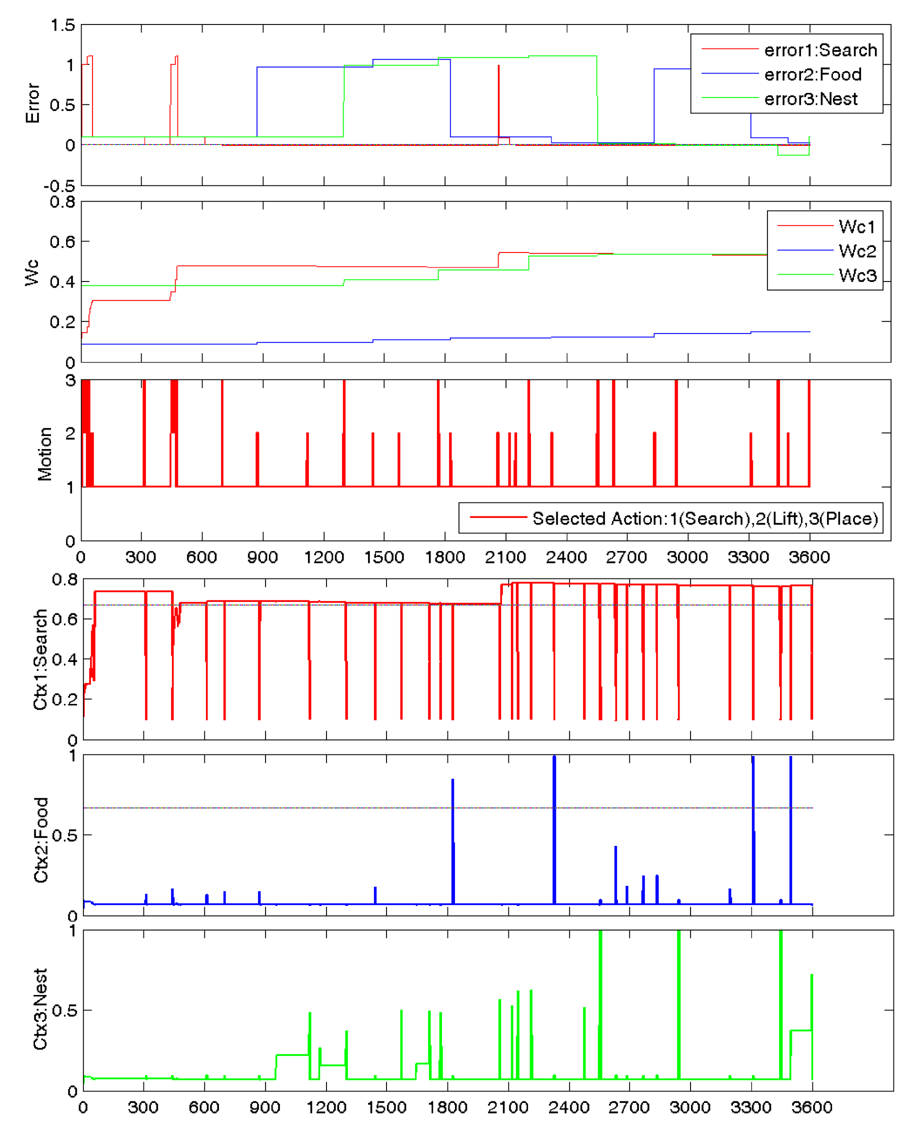
$$S(k) = Wc * I(k)$$
$$Ctx(k+1) = f(\lambda Ctx(k) + Thl(k) + S(k))$$
$$Str(k+1) = Wrf\big(Ctx(k)\big)$$
$$GPe(k+1) = f(-Str(k))$$
$$Stn(k+1) = f(Ctx(k) - GPe(k))$$
$$GPi(k+1) = f(WdStn(k) - Str(k))$$
$$Thl(k+1) = f(Ctx(k) - GPi(k))$$

Reinforcement Learning Rules

$$V(k+1) = Wv(k)S(k)$$
$$\delta c(k+1) = r_c + \mu V(k+1) - V(k)$$
$$Wv(k+1) = Wv(k) + \eta_c \delta c(k+1) S(k)$$
$$Wc(k+1) = Wc(k) + \eta_c \delta c(k+1) Ctx(k) S(k)$$

Test	# of steps	Wc11	Wc22	Wc33	Wr11	Wr22	Wr33	
Test1	3601	0.1201	0.875	0.3764	0.1	0.2641	0.1	Random normal initial parameter values
Test2	7077	0.1001	0.675	0.2243	0.1	0.2641	0.1	The values of Wc are decreased
Test3	869	0.1201	0.875	0.3764	0.9	0.2641	0.1	The value of Wr11 (Dopamine level) is increased
Test4	1041	0.5250	0.1503	0.6920	0.1	0.2641	0.1	Initials are set to the values after learning

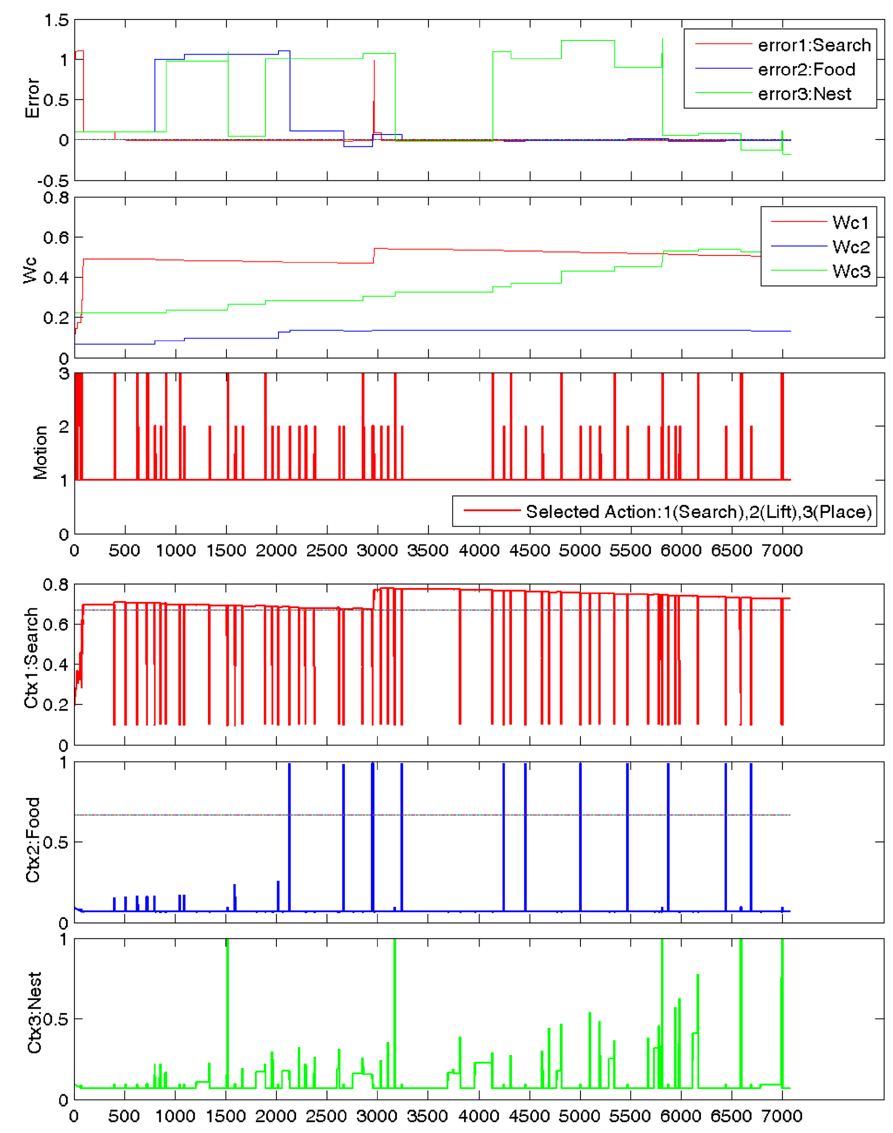
Sketches of Test 1



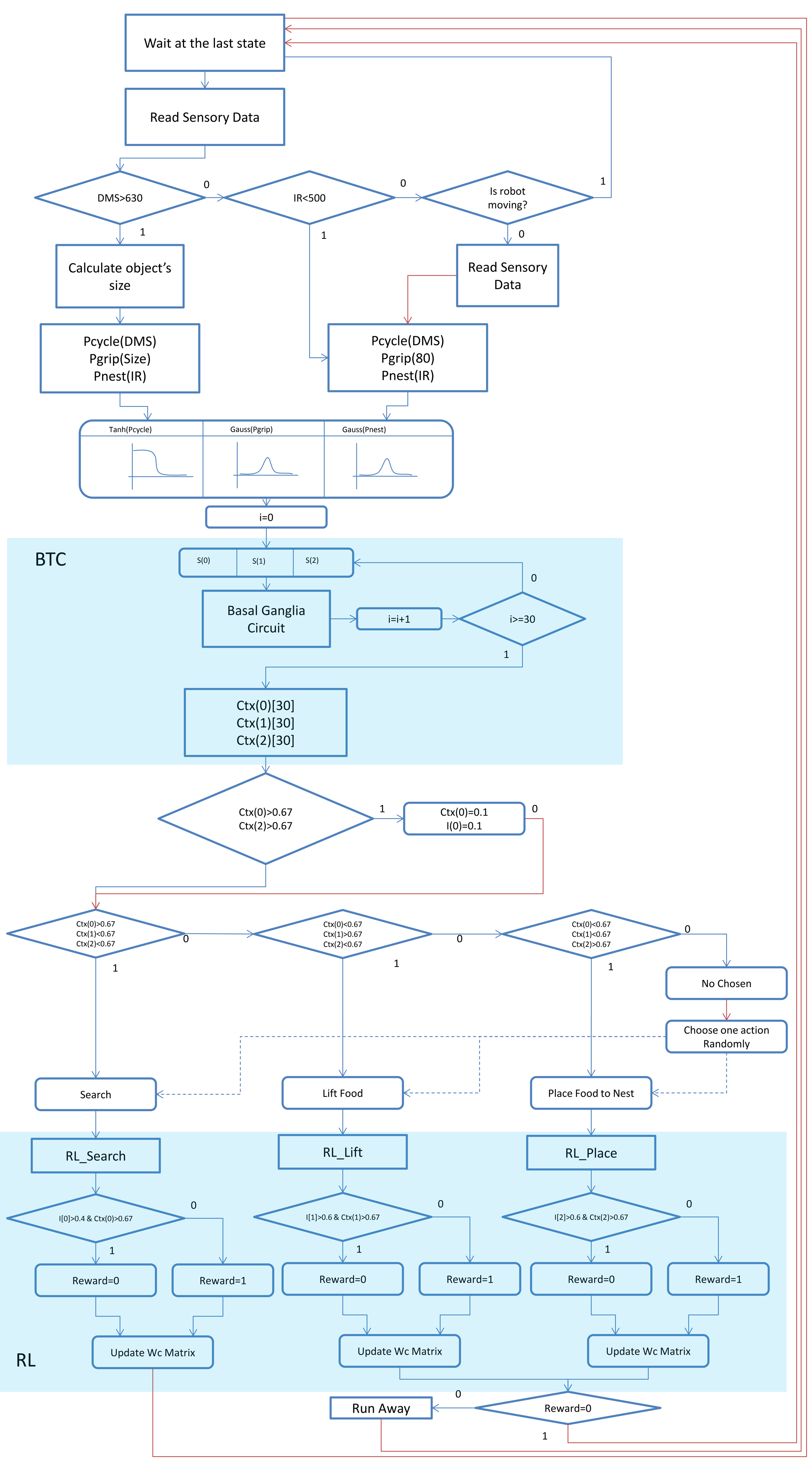
In Test 1, initials of Wc parameters, which help robot to comprehend environment, are randomly defined. After several trials robot learns to select the right action as error decreases to zero. Motion shows which action robot is realizing at that time. Ctx values shows that the action is selected if related Ctx value is over 0.67.

In Test 2, the initial values of Wc are decreased. Robot learns in this test. However, the learning time is longer and robot has to make more trials.

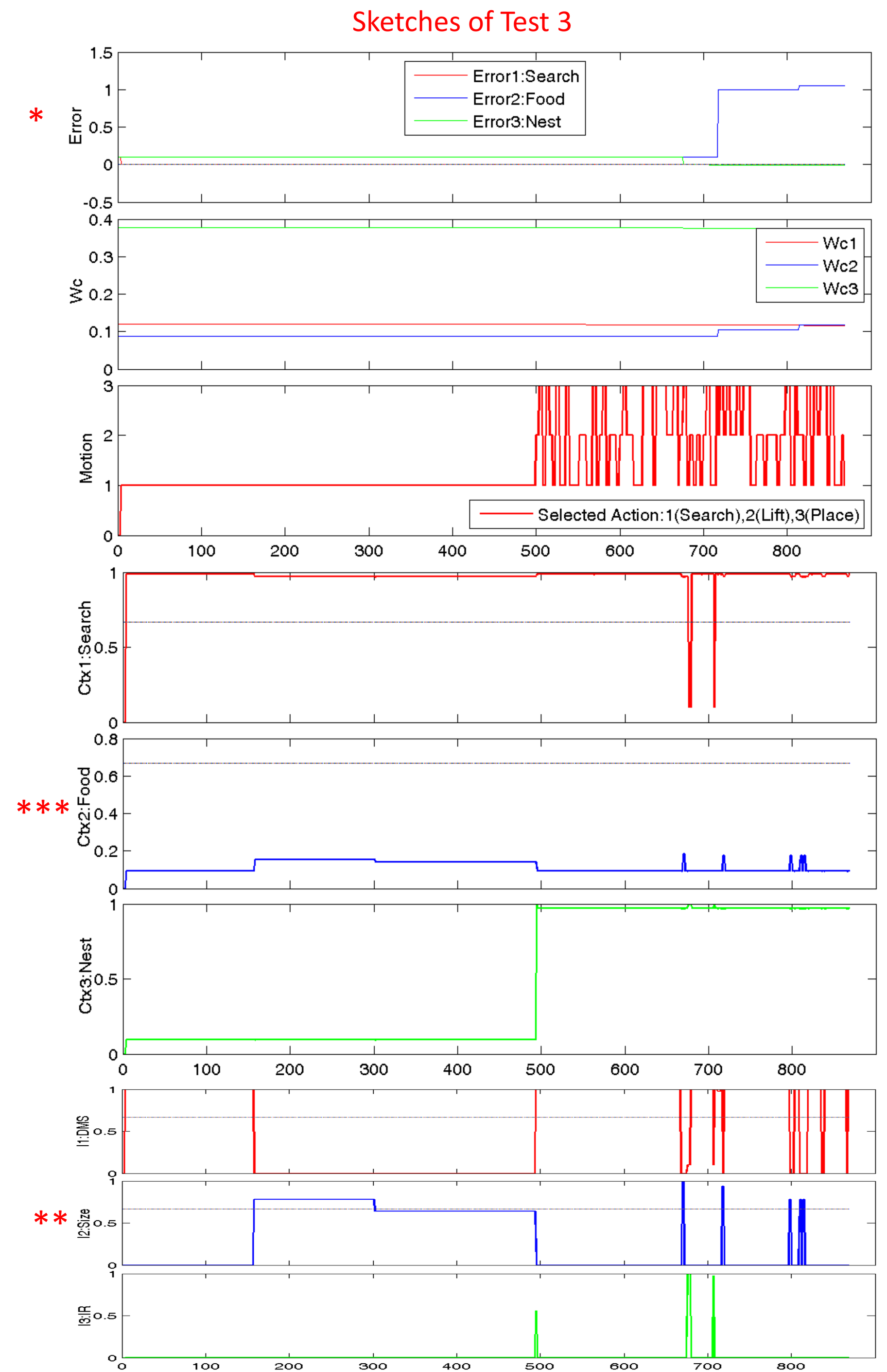
Sketches of Test 2



Flow Chart of the Robot's Programme



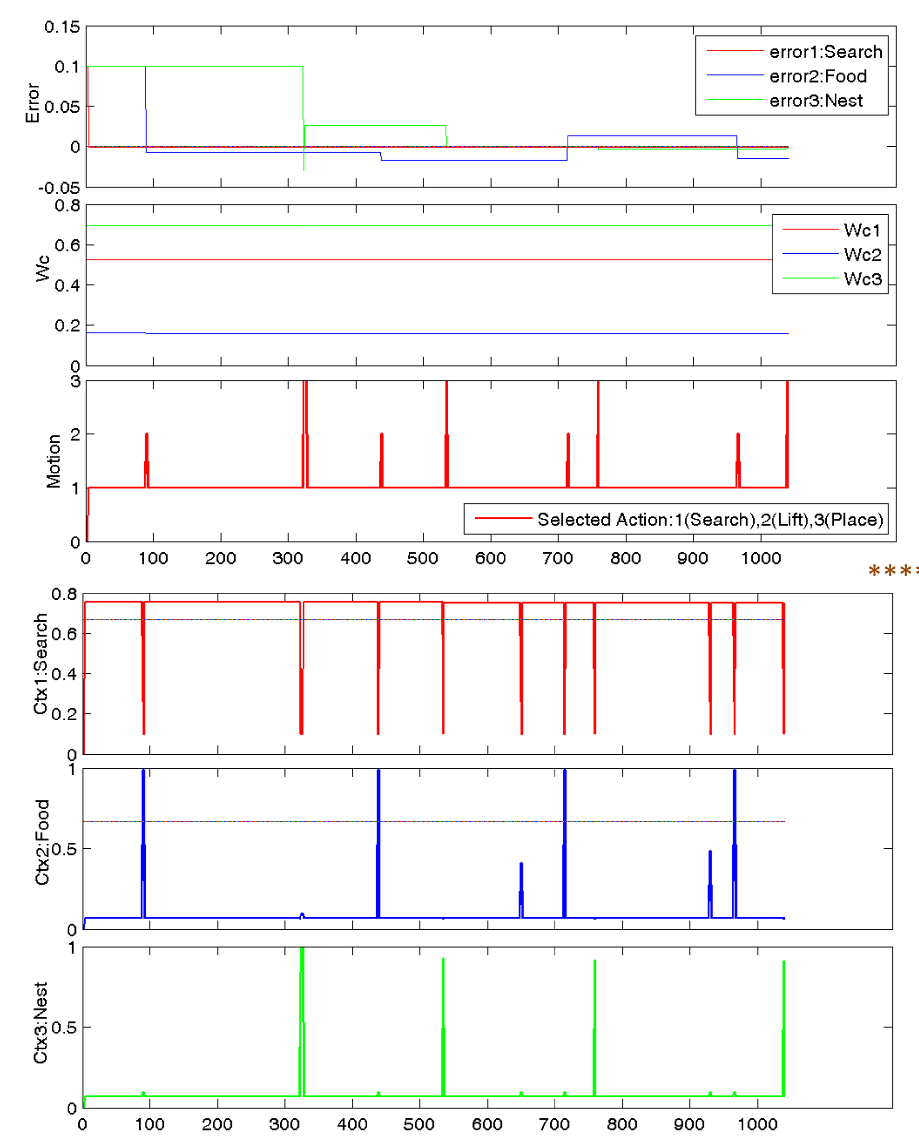
Sketches of Test 3



In Test 3, Wr, which represents dopamine level, is changed. Even the errors are low, robot cannot selects right action*. Robot encounters with a food after 150^{th} step, but it doesn't react to it because of Wr**. After encountering a nest at the 500^{th} step, robot tries to select two action at a time, which is not possible***. And it begins to oscillate between the actions.

In Test 4, the initials of Wc's are the values that after learning is accomplished. Robot selects the right actions without trials. Meanwhile, the motion sketch is parallel to the Ctx sketches****.

Sketches of Test 4



As conclusion, the model is compatible with a robot which has limited sensory properties. The robot is able to complete the foraging task depending on the parameters of the model.

References

*B. Denizdurduran, "Learning How To Select An Action: From Bifurcation Theory To The Brain Inspired Computational Model", İ.T.Ü. Fen Bilimleri Enstitüsü, Yüksek lisans tezi, 2012 http://www.simmag.itu.edu.tr/yayinlar/denizdurduran_Msc_Thesis.pdf

* Alexander, G.E., Crutcher, M.D, "Funcitonal architecure of basal ganglia circuits: neural substrates of parallel processing", Trends Neurosci., Vol. 13, No. 7,, 266-271, July 1990.

*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. http://www.simmag.itu.edu.tr/yayinlar/ref_4_sengor.pdf

*"Robotis e-Manuel", http://support.robotis.com/en/

*Prescott, T.J., Gonzalez, F.M.M., Gurney, K., Humphries, M.D. and Redgrave, P., 2006. A robot model of the basal ganglia: Behavior and intrinsic processing, Neural Networks, 19, 31–61.59

*Domjan, M. ve Grau J.W., 2003. The Principles of Learning and Behavior. 5th Edition., Thomson

*This work has been supported by Tübitak Project 111E264