Robotica: (2020) volume 38, pp. 1558–1575. © Cambridge University Press 2019
doi:10.1017/S0263574719001632

Experimental Study of Reinforcement Learning in Mobile Robots Through Spiking Architecture of Thalamo-Cortico-Thalamic Circuitry of Mammalian Brain

Vahid Azimirad* and Mohammad Fattahi Sani

Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
E-mail: m.fattahi93@ms.tabrizu.ac.ir

(Accepted October 19, 2019. First published online: November 18, 2019)

SUMMARY
In this paper, the behavioral learning of robots through spiking neural networks is studied in which the architecture of the network is based on the thalamo-cortico-thalamic circuitry of the mammalian brain. According to a variety of neurons, the Izhikevich model of single neuron is used for the representation of neuronal behaviors. One thousand and ninety spiking neurons are considered in the network. The spiking model of the proposed architecture is derived and prepared for the learning problem of robots. The reinforcement learning algorithm is based on spike-timing-dependent plasticity and dopamine release as a reward. It results in strengthening the synaptic weights of the neurons that are involved in the robot's proper performance. Sensory and motor neurons are placed in the thalamus and cortical module, respectively. The inputs of thalamo-cortico-thalamic circuitry are the signals related to distance of the target from robot, and the outputs are the velocities of actuators. The target attraction task is used as an example to validate the proposed method in which dopamine is released when the robot catches the target. Some simulation studies, as well as experimental implementation, are done on a mobile robot named Tabrizbot. Experimental studies illustrate that after successful learning, the meantime of catching target is decreased by about 36%. These prove that through the proposed method, thalamo-cortical structure could be trained successfully to learn to perform various robotic tasks.

KEYWORDS: Reinforcement learning; Spiking neural networks; Mobile robot; Thalamo-cortico-thalamic circuitry; Dopamine modulator.

1. Introduction

Learning procedure in mammalian brain architecture is one of the most fascinating phenomena in the world, which is a result of evolution. Unraveling the mystery of high-level learning procedure in the brain could be promising for online learning of robots, specifically in unknown environments. A highlighted section of mammals' brain named "thalamo-cortical" circuitry (which contains circuits between thalamus and cortex) is a reputed architecture for studying complicated functions of the brain such as learning.[1] Some researchers attributed the fast learning ability of the brain to neural circuits between thalamus and cortex,[2] and others investigated action learning and controlled actions process through thalamo-cortical circuits.[3] To study the possibility of learning in thalamo-cortical circuitry, a game world is utilized.[4] One of the major responsibilities of thalamo-cortico-thalamic

* Corresponding author. E-mail: Azimirad@tabrizu.ac.ir

(TCT) circuitry is to involve in reinforcement learning.[5,6] It studies how biological systems can solve instrumental conditioning problems.[7] Recent researches suggest that learning and instrumental conditioning are integrated into a network centered on thalamus.[8] Besides, thalamo-cortical circuitry is connected to the motor system, so it may be the best model for motor learning of robots through instrumental conditioning.[2,9] To apply instrumental conditioning on the detailed structure of thalamo-cortical circuitry, the spiking model of neurons is needed. Although traditional artificial neural networks are accep tools for supervised learning, they do not show reasonable results when there is not any desired output (as we face in reinforcement learning problems). Spiking neural networks (SNNs) are new, powerful, and biologically plausible tools for behavioral modeling of natural neurons.[10,11] They are gaining much attention due to their biological basis, and recent developments in computational neuroscience generated infrastructure for functional study of brain based on the SNNs.[12] There are various models of spiking neurons (e.g., Integrate and Fire, Izhikevich, and Hodgkin–Huxley), and they can be utilized based on their model complexity and computational cost.[10] Izhikevich et al. studied the large-scale model of thalamo-cortical structure of the mammalian brain[13] and proposed an algorithm for instrumental conditioning based on the SNNs.[14] Although SNN-based models of thalamo-cortical system are widely discussed in neuroscience[15] and there are some applications of them on robot learning systems,[16–21] this subject still needs more work done.

Robotic learning is a multidisciplinary field between computer science, neuroscience, and engineering. Some methods of reinforcement learning are applied to robotic systems, but they are not biologically plausible (e.g., value function[17] and temporal difference[18] methods). Among all works in robotic learning, there are little studies about SNNs. However, applications of them in the learning of artificial systems and robots are growing.[19] Some researchers implemented SNNs to control iCub humanoid robot (iCub),[20] DARwIn-OP humanoid robot,[16] and Khepera I robot,[21] but they used only simplified two-layered sensory–motor structure. Some others studied action selection learning based on basal ganglia.[5] A simplified model of thalamo-cortical circuits to use in a reinforcement learning system is studied in ref. [5], but its application in robotics is not considered, and the model is not biologically plausible as they used a simplified model of neurons.[22] In other works on humanoid robots, amygdala-thalamo-cortical structure is used,[23] but they did not study reinforcement learning. Studying the reinforcement learning procedure considering its biological details would be a more realistic approach to control robots. I shows a brief study of related works in this area.

As it is shown in Table I, Leaky Integrate and Fire (LIF) is a dominant model in previous works, but accurate functions of single neuron cannot be achieved through it.[5] Besides, there are just a few experimental implementations compared to simulation studies. Some studies were conducted to implement the learning procedure on autonomous robots using SNNs.[28] However, no one addressed the learning problem in autonomous robots using a complete model of thalamo-cortical circuits. Recently, the functionality of excitatory and inhibitory neurons in the learning of a single joint robot is discussed in ref. [53]. But it is done through a simple sensory–motor structure of the cortex. Through simulating thalamo-cortical dynamics, one may include high-level functions of the mammalian brain into the learning model of robots.

The main aim of this paper is to study the functionality of thalamo-cortical structure of the mammalian brain in the learning of mobile robots. It sheds new light on biological-based robotic learning by implementing TCT architecture of mammalians' brain on robots. The application of thalamo-cortical structure in robotic learning is a new topic. The main contribution of this paper is to study reinforcement learning on mobile robots in which the spiking neuronal model of TCT circuitry is used. Dopamine-modulated spike-timing-dependent plasticity (STDP) is used to incorporate reinforcement learning on the model.[14] Also, a new formula for releasing dopamine is proposed based on the robots' environmental inputs. In other words, reward delivery in the nervous system is related to external conditions instead of internal neural pathways, which is not discussed in any of the previous works about SNNs. Among all previous studies in this area, there is no work to study the reinforcement learning of mobile robots through the spiking structure of the thalamo-cortical system of the mammalian brain. The rest of this paper is organized as follows: in Section 2, SNN model of a single neuron, unsupervised learning, STDP, dopamine-modulated learning of robots, and a new architecture of thalamo-cortical connectivity are explained. Simulation and experimental studies are presented in the third section to show the effectiveness of the proposed method. Results and discussion are addressed in the fourth section, and finally, Section 5 concludes this paper.

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table I. Details of recent works on SNN based learning of robots.

Ref No.	Neuron model	Network architecture	Experimental test bed	Simulation studies
24	Izhikevich	Thalamo-Cortico-Thalamic (TCT)	Sleep Awakefullness, SpiNNaker	NO
25	Functional	TCT	NO	Alpha rythms in Alzaimer
13	Izhikevich	TC	NO	Brain simulation
26	Izhikevich and LIF	Cerebral cortex, thalamus and amygdala	General-Purpose Graphics Processing Units	NO
27	LIF	BG	NO	Action Selection and Oscillatory Activity
21	LIF	BG	Khepera^{TM} I (MR), Action Selection	NO
20	LIF	BG	Problem solving, iCub +5DOF arm	NO
5	LIF	BG	NO	Action selection
15	Izhikevich	BG & SNN	NO	Neurological
28	LIF	BG	NO	Neurological
3, 29	LIF	Cortico-basal ganglia	NO	Action Learning
16	Functional	BG, thalamus and cortex	DARwIn-OP humanoid, Action selection	NO
30	Izhikevich	3-layer	NO	TF (MR)
31	Izhikevich	*	TriBot (MR), OA, TD	NO
32	Izhikevich	FFN	iCub	NO
33	Izhikevich	*	NO	Maze PP (MR)
34	Izhikevich	*	NO	TF (MR)+Arm
35	LIF	*	NO	Olivier Michel's Khepera Simulator, WF
36	LIF	DCS, SFCS	(MR), TF, OA	NO
37	LIF	*	Virtual (MR)	
38	LIF	Insect inspired model	NO	TF, NAV
39	LIF	Self-organized network	NO	Pioneer-3 WF (MR)
40	LIF	*	Virtual insect in Digital CMOS	
12	LIF	*	NO	Memristor + (MR) OA
41	SRM	Retina model	MObiMac (MR), TF	NO
42	SRM	DCS	(MR), WF	NO
43	SRM	ASNN	Khepera-I OA	NO
44	SRM	*	Bioloid Robot walking motion	Virtual Environment
45	SRM	3-layered FSNN	2-Dof manipulator	NO
46	IF	*	NO	A mobile robot built by Institute of Automation, Chinese Academy of Sciences (CASIA) (MR)
47	IF	*	2 gripper & beaglebone black	NO
48	IF	DCS	NO	Corridor-Scene Classification

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table I. Continued.

Ref No.	Neuron model	Network architecture	Experimental test bed	Simulation studies
49, 50	IF	*	NO	WF,OA,TF (MR)
51	Adaptive IF	Fruit fly olfactory inSpired model	TF (MR)	NO
52	SKAN	*	Maze solving (MR), Field-Programmable Gate Array	NO

Neuron model acronyms: Spike Response Model (SRM), Integrate-and-Fire (IF), Synaptodendritic Kernel Adapting Neuron model (SKAN)

Implementation acronyms: Mobile Robot (MR)

Architecture acronyms: Basal Ganglia (BG), Thalamo-Cortico-Thalamic (TCT), Thalamo-Cortical (TC), Fuzzy Spiking Neural Network (FSNN), Delay Coding for Sensory input (DCS), Spike Frequency Coding for sensory input (SFCS), Aplysia-like Spiking Neural Network (ASNN), Feed Forward Network (FFN)

Task acronyms: Wall Following (WF), Obstacle Avoidance (OA), Target Detection (TD), Path Planning (PP), Target Following (TF), Navigation (NAV)

*Unknown

2. Methods and Materials

Here, neuronal modeling is presented, reinforcement learning as well as SNNs is discussed, and then, the new structure is presented.

2.1. Mathematical representation of a single spiking neuron

Neurons transmit electrical and chemical signals to send information. In the simplified model of a single neuron, a tube-like compartment named axon is considered to conduct electrical impulses (spikes) away from the cell body. The junction between two terminals of neurons is called synapse. Here, the electrical properties of signal transmitting are used because of simplicity. Neuronal models are defined in the form of Ordinary Differential Equations. Among all the proposed models in the literature (such as the easily explained model of LIF and complex model of Hodgkin–Huxley), simple and accurate models such as the Izhikevich are better for this work because of high accuracy in covering the properties of neurons and low computational cost. Here, the Izhikevich differential equations of the single neuron are used for modeling the architecture of the TCT system.[54] The dynamic equations of the single neuron are shown in Eqs. (1), (2), and (3):

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I \tag{1}$$

$$\frac{du}{dt} = a\,(bv - u) \tag{2}$$

$$\text{if } v \geq 30\,\text{mV},\ \text{then} \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{3}$$

in which v and u represent the membrane potential and the membrane recovery parameter, respectively. a, b, c, and d are the dimensionless parameters and set according to physiological data.[54]

2.2. SNNs and learning procedure

SNNs are widely used in neuroscience, but their application in artificial intelligence and reinforcement learning is not discussed well. Here, scaling down the mammals learning procedure comprises sensors, processing systems, and motors sections as well as dynamic modeling of neuronal behavior in terms of STDP and rewarding. The processing system contains N artificial spiking neurons that are connected. It is assumed to have some excitatory and inhibitory layers. Figure 1 shows a simplified neural model of the proposed design.

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

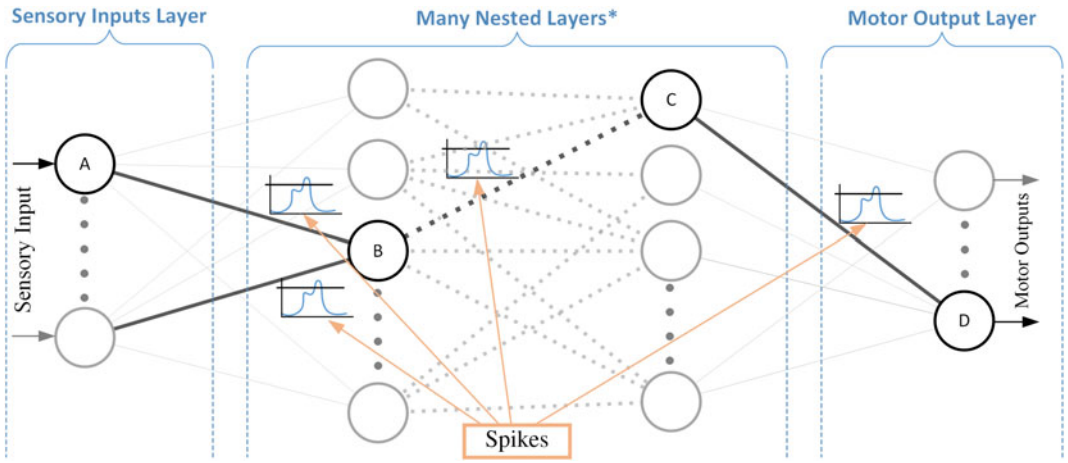

Fig. 1. Simplified neural model of spiking neural networks.
* It may contain lots of excitatory and inhibitory layers.

As it is shown in Fig. 1, SNNs contain sensory neurons, some interneurons (INs), and motor neurons. Sensory data of environment enter to network as spikes, and they continue their way through different parts of the network toward the output parts (motor neurons). Spikes are emitted according to the connection map of neurons and using Eqs. (1), (2), and (3). If spikes are emitted to some neurons in the motor layer and make them fire, there will be a movement for the actuator that is connected to those motor neurons and the amount of movement is proportional to the number of the fired neurons at that part. This spike-based system is used as a platform for the proposed learning algorithm.

It is supposed that specific action of the body in mammals is because of activating some specific pathways between sensory and motor areas in the central nervous system. When the right action (e.g., catching the target) happens, the concentration of dopamine will be increased and synaptic weight reinforcement of neuronal chains (neuronal pathways) that are involved in that pathway will happen. In other words, proper movements of the body will result in giving rewards and it will strengthen the neural pathways that are involved with this action. For example, most dopamine neurons show activation after food rewards and conditioned stimuli.[55] It is supposed that a robot, as a bio-inspired system, has sensors, actuators, and processing systems. Sensors of the robot are activated when there is a target in the detection range of them, and they send information regarding the distance of the target from the robot. Hence, when a specific sensor is activated, the reinforced pathways of neurons that result in activation of specific motor neurons are generated and some causal processes that lead to catching the target in less time are made. Figure 2 shows the flowchart of applying reinforcement learning on the proposed system. It iterates until to find targets for 100 times.

As it is shown in Fig. 2, there are two sources of input for the neural network: informational current input from sensory cells and a random current input named $I_{motorbabbling}$. Informational current input from sensory cells contains electrical current in which its intensity is analogous to the inverse of the distance of the target from the robot. Sensory data regarding the inverse of the distance of the target from the robot enter into the neural network as electrical current every 16 ms. Also, some random electrical current is considered to enter into each right and left of the section in the motor area to make random movements named motor babbling which is called $I_{motorbabbling}$. The output of the neural network are the spikes of motor neurons which result in the movement of the robot in the environment through motors and wheels. The specific action of the body is the result of synchronized activating of some specific neurons (neuronal pathways) between sensory and motor areas in the nervous system. When the right action (e.g., reaching the target) happens, the reward will be given to the system and the concentration of dopamine will be increased. This will affect the synaptic weights of firing neurons. In other words, when dopamine is released, neurons that are involved causally in those pathways will be reinforced and synaptic weights of active neurons that are not involved in causal firings of other active neurons will be decreased. Here, the learning process consists of two stages: unsupervised and reinforcement learning. Here, the unsupervised learning stage is modeled based on

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

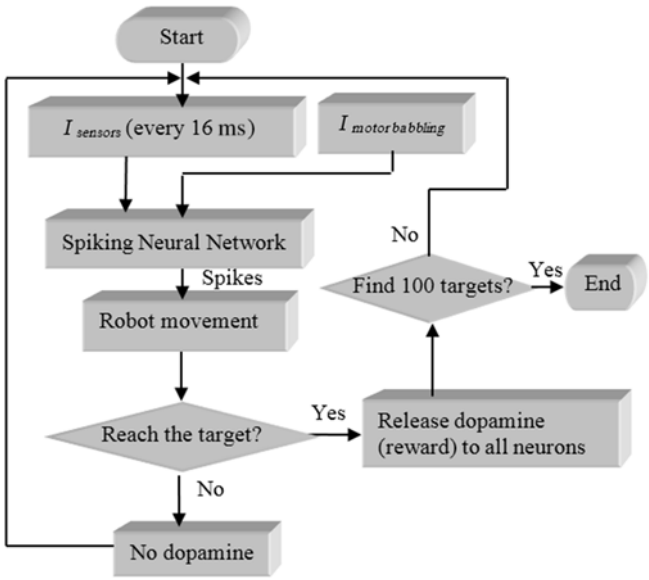

Fig. 2. Flowchart of applying reinforcement learning on the proposed system.

STDP. Unsupervised learning alters the weights of synaptic connections between neurons based on the spike times of pre- and postsynaptic neurons. Assume two arbitrary neurons A and B in Fig. 1 which are connected by a synapse so that A is a presynaptic neuron and B is a postsynaptic neuron. If neuron B fires before the firing of neuron A, it means that they have no causal relationship and the weight of the synaptic connection between A and B will be decreased, and vice versa. However, the rate of decrease and increase in weight of synapse depends on the temporal difference of firings. Larger time gaps between firings of neurons A and B lead to less variation of synaptic weight. The reason behind STDP is to magnify the synapses of neurons that are firing together (there is a causal relationship between them) and to weaken the synapses of neurons that have no causal relationship. The overall strength of synapses (S) determines the rate of electrical current through neural pathways (Eq. 4), and it is modified during learning process according to the value of eligibility trace q and dopamine d (Eq. 5).[14]

$$\dot{I} = S \tag{4}$$

$$\dot{S} = q.d \tag{5}$$

$$\dot{q} = -\frac{q}{\tau_q} + STDP(\tau_q)\delta(t - t_{pre/post}) \tag{6}$$

$$\dot{d} = -\frac{d}{\tau_d} + dop \tag{7}$$

where I represents the amount of electrical current across each synapse, S shows the strength of synapse, q is the eligibility trace, STDP is the STDP, τ shows the decay parameters of STDP, d is the concentration of extracellular dopamine in synaptic junction, and dop shows the release of dopamine in neural networks when external reward is given to robot. There is a time-dependent variable (q) for each firing neuron in STDP algorithm, which is decaying by time. Assume that an arbitrary neuron A in Fig. 1 has just spiked now. The synapses of all postsynaptic neurons of fired neuron A are decreased by the value of q, and the synapses of all presynaptic neurons are increased by the value of q. Therefore, if a presynaptic neuron had spiked recently, the connection between that neuron and mentioned neuron A will be strengthening (e.g., the connection between the mentioned neuron and a postsynaptic neuron, which fires before that, is decreased by the value of q). If the postsynaptic neuron fired a long time ago, the value of q will be much smaller and the decrements in synapses weight will not be significant. The coefficients of magnifying and descending synapses are important parameters in balancing the Spiking Neural Network (SNN)s. These parameters should be tuned

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table II. Values of a, b, c, and d parameters for neurons in the TCT structure.

Neuron name	Neuron type	a	b	c	d
TCR_R	RS	0.02	0.2	−65	8
TCR_L	RS	0.02	0.2	−65	8
PY_R	RS	0.02	0.2	−65	8
PY_L	RS	0.02	0.2	−65	8
$Ex.-IN$	RS	0.02	0.2	−65	8
IN	FS	0.1	0.2	−65	2
TRN	IB	0.02	0.2	−65	4
$Fast-IN$	FS	0.1	0.2	−65	2
$Slow-IN$	LTS	0.02	0.25	−65	2

before the beginning of the reinforcement learning process, so that after some time of running, the value of average synaptic weights of all neurons will be the same as it is the initial value[13] (If the coefficient of magnifying is greater than the coefficient of descending, the average of all synaptic weights of all neurons grows and the network will not be balanced). Formulation of STDP is used beside reinforcement learning which is based on the dopamine release during reward.[14]

The reinforcement learning mechanism is based on the release of dopamine. In other words, the delivery of reward is simulated as an activity of dopaminergic cells, which increases the concentration of dopamine in the synaptic junction. Eqs.(5) and (7) illustrate how dopamine-modulated STDP addresses reinforcement learning on the synaptic level. Step increasing the concentration of dopamine (as Eq. 7), which occurs because of catching the target, results in faster synaptic changes and magnifying synapses that are incorporated in the generation of right neural pathways to preform current action of the body.[14]

2.3. TCT architecture of the SNN

Why we use the thalamo-cortical structure? Thalamo-cortical pathways play an important role in high-level behaviors of the sensory–motor system in mammalian brains.[13] In the literature, some previous works utilized cortical sensory–motor architecture, which is very simple.[56,57] Some other works are about the implementation of TCT on Spinnaker to study the function of the brain in sleep states,[25] which is based on the proposed model of ref. [13]. However, they used the "neural mass model",[58] which is not realistic because of its merely functional point of view. In our work, an improved architecture of thalamo-cortical circuit[25] is used. The cortical module comprised of the excitatory pyramidal (PY) cells, which are divided into right and left sections (PY-R and PY-L), the excitatory INs, slow inhibitory INs, and fast inhibitory INs. The thalamic module consists of the right and left thalamic relay cells (TCR-R and TCR-L), INs, and thalamic reticular nucleus (TRN). TCRs have an excitatory role, while IN and TRN cells have the inhibitory role. Figure 3(b) shows the connectivity map of subsystems, and Table II shows the values of a, b, c, and d parameters for different neuron types of each section according to the Izhikevich model.

Figure 3(b) shows the connectivity map of sub systems.

As it is shown in Fig. 3(b), PY and TCR sections are divided into two subsections because of two sensors and two motors; in addition, it is inspired by two hemispheres of the mammalian brain. The details of connections between neurons of each part are demonstrated in Table III.

The first column in Table III demonstrates the type of neurons in each part of the TCT architecture. In Table III, "1" means that there is a connection from the module in the row to the module in the column, while "0" means there is no connection between them. The number of synapses of each connection is available in the second row of Table III. A full model of thalamo-cortical system is presented in ref. [13], but it has more computational cost than the current model. More recently, a simplified version of this structure is discussed in ref. [24], in which TCT architecture based on the Izchevich model of a single neuron is implemented on Spinnaker to study sleep and wakefulness cycles of the brain. Here, the connectivity details in Table III are based on ref. [24]. The number of neurons in the cortex module is also scaled down as it was in ref. [13].

The proposed learning system with TCT architecture is tested on a robot. Sensory data, as a spike of neurons, come into the thalamic system through electrical current, which results in the spiking

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table III. Detailed parameters of the architecture.

From\to				**TCR_R**	**TCR_L**	**PY_R**	**PY_L**	**Ex.IN**	**IN**	**TRN**		**Fast IN**		**Slow IN**
Number of synapses				**6**	**6**	**75**	**75**	**22**	**2**	**6**	**6**	**13**	**13**	**12**
	Neuron index	**No. of neurons**	**Type of neurons**											
TCR_R	0–25	25	RS	0	0	1	1	0	0	1	0	0	0	0
TCR_L	26–50	25	RS	0	0	1	1	0	0	1	0	0	0	0
PY_R	51–350	300	RS	1	1	0	0	1	1	1	0	1	1	1
PY_L	351–650	300	RS	1	1	0	0	1	1	1	0	1	1	1
Ex.IN	651–830	180	RS	0	0	1	1	0	0	0	0	0	0	0
IN	831–840	10	FS	1	1	0	0	0	1	0	0	0	0	0
TRN	841–890	50	IB	1	1	0	0	0	0	1	1	0	0	0
Fast-IN	890–1000	110	FS	0	0	1	1	0	0	0	0	0	0	0
Slow-IN	1001–1090	90	LTS	0	0	1	1	0	0	0	0	0	1	0

FS, fast spiking; RS, regular spiking; IB, intrinsically bursting; LTS, low-threshold spiking. “1” means there is connection between two parts. For Example, the first row of the table shows that the TCR_R section is connected to PY_R, PY_L, and TRN. There are 25 RS neurons in the TCR_R part. The second row in the table shows the number of synapses for each part. For instance, there are 75 synapses from TCR_R to PY_R.

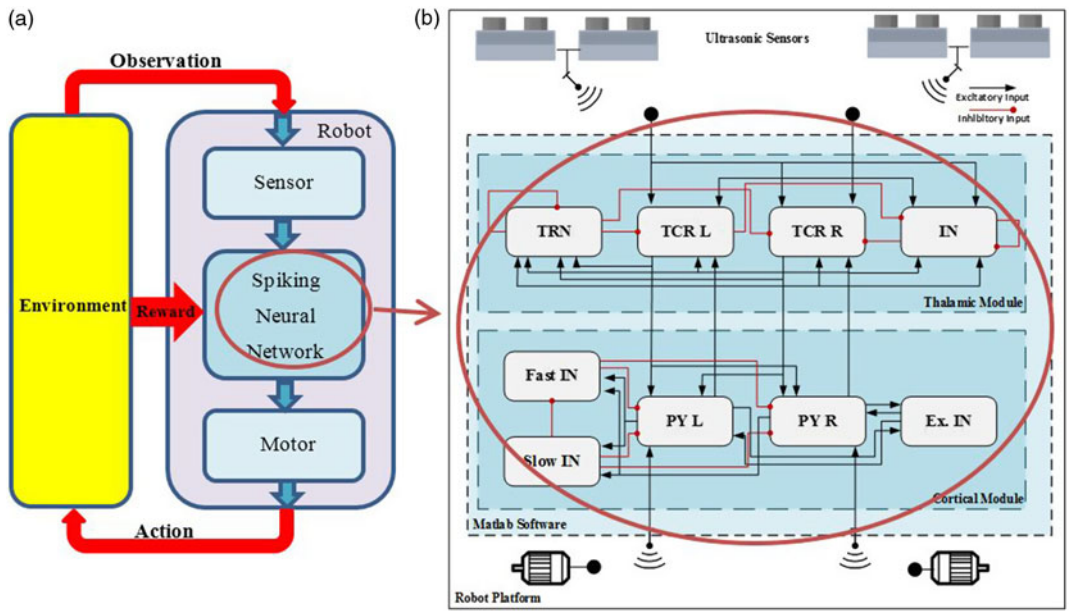

Fig. 3. (a) Schematics of reinforcement learning system. (b) TCT architecture used in this study.

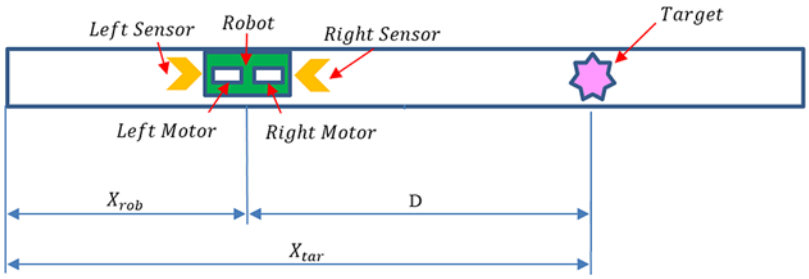

Fig. 4. Schematic of the test setup comprised of a mobile robot and a target.

of the related neurons. According to the connection map of TCT circuitry, spiking of presynaptic neurons in SNNs may cause to the spiking of postsynaptic neurons. So, if synaptic weight is large enough to make membrane voltage of neuron reach to action potential threshold, propagation of spikes in the network will be continued until the firing of motor neurons. Besides, there are some random electrical currents for motor neurons to simulate motor babbling in the central nervous system of mammals. Excitation of any specific actuator of the robot is related to the rate of spikes of neurons in each specific motor subsection.

3. Simulation and Experimental Studies

Suppose that a linear mobile robot has a neural network structure shown in Fig. 3 alongside its parameters in Table III and there is a target in environment. Target foraging task is chosen to implement the proposed learning architecture. When the robot moves and catches the target, the reward will be given to the robot as a step increase of dopamine in SNN.

3.1. Simulation studies

This test aims to train the robot to move towards the target and catch it. When the robot catches the target, another target will appear randomly. The initial position of both robot and target is assumed to be in the context of the line (limitation of movement is 100*cm*) and they are randomly generated to produce equal probabilities of detection for left and right sensors (Fig. 4).

As it is shown in Fig. 4, the test setup includes a 1-degree-of-freedom (DOF) linear mobile robot equipped with two proximity sensors on its right and left sides, and a target. *D* shows the distance of

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

the target from the robot which is detected by sensors. When we start the test, the robot will move to the left and right sides randomly according to random spikes of neurons in the motor section (PY neurons) which is called motor babbling inputs. In this stage, there is no clear pathway from the input sensors to the motors so the robot is somehow blind. When the target is at the right or left side of the robot, right or left sensory neurons in "Thalamic Relay Cells (TCR)" in section TCR_R or TCR_L receive normally distributed input current from the right and left sensor, respectively. Distance signals from sensors enter the neural network every 16 ms and the mean value of this normally distributed electrical current is analogous to distance inverse of the target from the robot (Eqs. 8 and 9). Simulated program is running in the frequency of 4 kHz.

$$\text{if } D_R < Th_{sensor} \longrightarrow I_{TCR_R} = poiss\_rand \ \ (\alpha) \tag{8}$$

$$\text{if } D_L < Th_{sensor} \longrightarrow I_{TCR_L} = poiss\_rand \ \ (\beta) \tag{9}$$

Where D_R and D_L show distance of the target from the robot according to Fig. 4. Th_{sensor} is the detection range of the sensor, α and β are tunable parameters (which are related to distance of the target from robot), I_{TCR_R} and I_{TCR_L} are right and left TCR cells input currents, respectively. If the target is at the right side of the robot and in the detection range of robots, D_R will be a positive number and D_R will be zero. On the other hand, when the target is at the left side of the robot and in the detection range of the sensor, D_L will be a positive number and D_R will be zero. So, spikes will be emitted through the network according to Eqs. (1), (2), and (3) and synaptic weights of neurons will be affected and altered according to Eqs. (4), (5), and (6) and connection map of neurons. Propagation of spikes is continued to affect motor neurons of a cortical column. The difference between the number of spikes in right and left PY cortical neurons sections is assumed to determine the direction and velocity of the robot (as Eq. 10).

$$Vel = SP_{PY_R} - SP_{PY_L} \tag{10}$$

Where *Vel* is the velocity of robot, SP_{PY_R} and SP_{PY_L} are the number of spikes in right and left PY (PY motor neurons of cortex), respectively, which are calculated every 16 ms. If *Vel* is positive, the robot will moves to the right, otherwise, it will move to left. The rewarding system is based on the dopamine release. If the robot happens to catch the target randomly, dopamine will be released and will affect all neurons of the network according to Eqs. (7) and (11). In other words, it accelerates the strengthening of synapses that have a positive amount of STDP and weakening synapses that have negative STDP value.

$$\text{if} D_i < \delta \longrightarrow dop = \frac{\gamma}{tms} \tag{11}$$

Where D_i is the distance of robot from the target in ith step, δ and γ are tunable parameters, *dop* is the amount of Dopamine that will be released and *tms* is the number of iterations. Dividing the value of dopamine into the No. of iterations results in a more effective learning process according to the instrumental conditioning methods.[14] Simulation is run for 12000 s and data is recorded to evaluate the results. In simulation studies, δ and γ are taken to be 0.007 m and 1, respectively. The sensor range value is taken 100 cm and the sampling time of sensors and motors are set to be 16 ms.

3.2. Experimental studies

The trained network is implemented on a mobile robot named Tabrizbot. It is a mobile robot equipped with ultrasonic proximity sensors on its left and right side and has Wi-Fi communication as well as an onboard liquid crystal display, which makes it suitable to run the program and debug it. The robot has two DC-Motors connected to two active wheels and there are also two passive wheels to prevent tipping over. Figure 5(a) shows functional block diagram of the robot.

As is seen in Fig. 5, the proposed algorithm is executed in a mathematical software on the PC and communicates with the robot through the Wi-Fi connection and TCP/IP protocol using ESP 8266 Wi-Fi Module. LPC1768 (ARM CORTEX-M3) Microcontroller with real-time operating system (RTOS) programming is used in the Tabrizbot to achieve maximum time optimization. The frequency of communication between PC and microcontroller is set 10 Hz. Right and left ultrasonic sensors (SRF-04) continuously send distance data to the main program. On the other hand, the computer runs the

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table IV. Parameters of the Tabrizbot robot.

Parameter	Value	Unit
Distance between two wheels	15	cm
Mass of robot	0.768	kg
Mass of wheel	0.033	kg
I robot	3.22908e-3	kg m^2
I wheel	1.03978e-6	kg m^2
Ultrasonic sensors measuring range	80	cm
Motors RPM	560	RPM

RPM: rotation per minute
I: moment of inertia

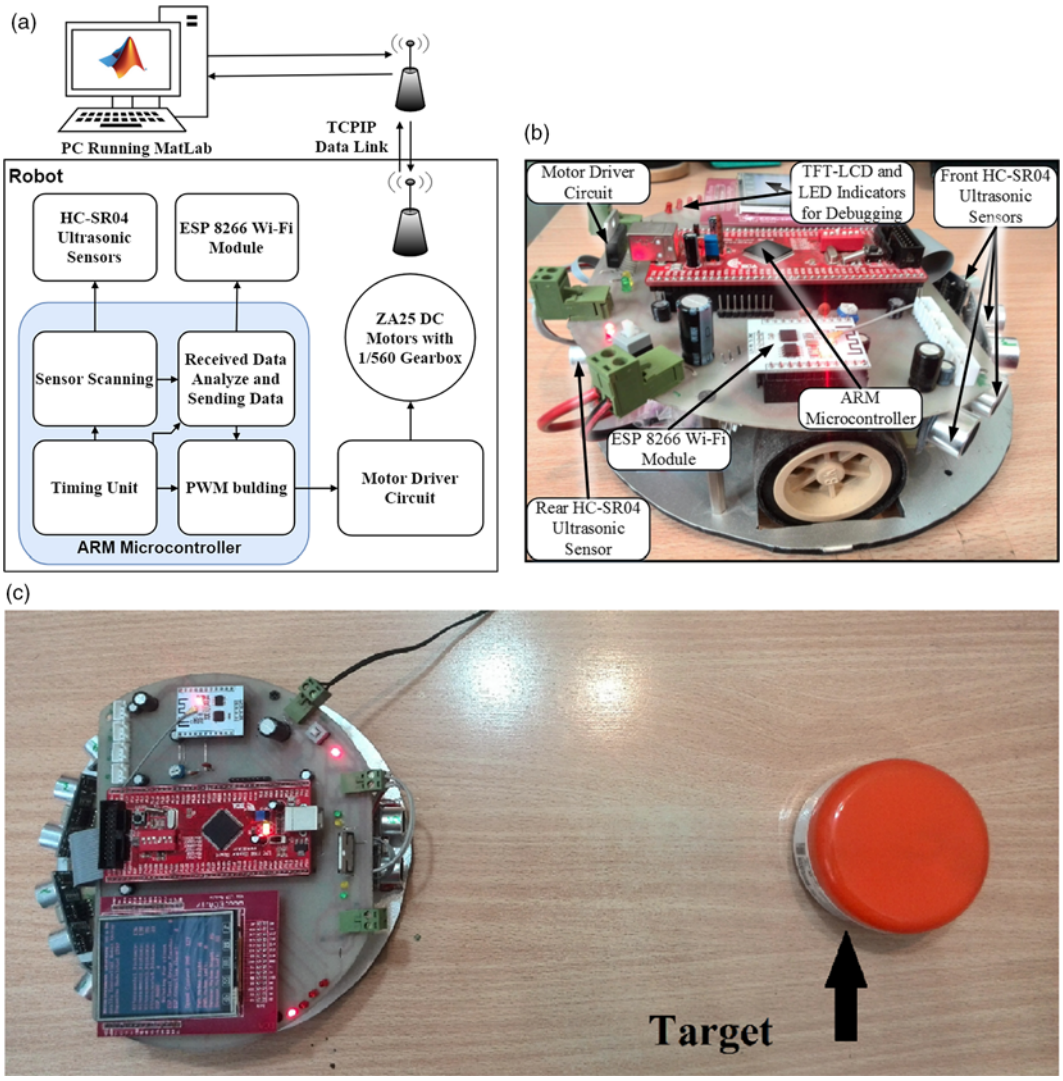

Fig. 5. (a, b) Tabrizbot and its functional block diagram. (c) Target attraction experiment.

SNNs and the value of the velocity of actuators are calculated at each iteration according to Eq. (10). Velocity parameters are sent back to the mobile robot. After decoding data, the microcontroller in the robot drives the motors with the appropriate pulse width modulation values. The characteristics values of the Tabrizbot robot are shown in Table IV.

In each experiment, as it is shown in Fig. 5, a target is placed on one side of the robot in random distances and the robot is expected to eventually succeed to catch it. Pseudocode of reading the sensors data and flowchart of implementing code for experimental setup is shown in the Appendix.

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

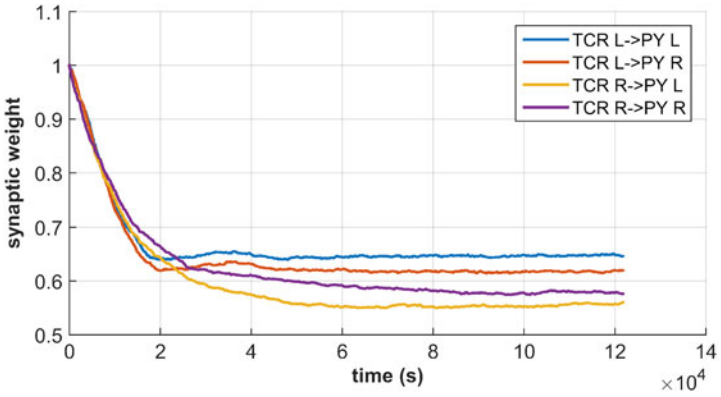

Fig. 6. Average weight of synapses of thalamic neurons to cortical neurons.

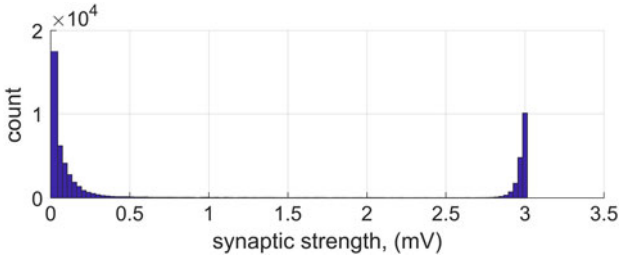

Fig. 7. The distribution of all synaptic weights in the network.

4. Results and Discussion

Simulation studies showed that the robot learns how to catch the target and some experimental tests are done to show the effectiveness of the proposed method. Experiments are repeated 10 times, and the behavior of robot before learning and after learning is compared in terms of foraging time and average synaptic weights between thalamic and cortical modules. The mean time of catching the target for the untrained system is 22.1, and the standard deviation is 14.8 s, while these values for the trained system are 14.3 and 13.4 s, respectively. Application of Student's paired t-test on two series of data showed that probability associated with a two-tailed distribution is less than 0.25. Figure 6 shows the average weight of synapses in the thalamic system to the cortical section of the proposed architecture.

Figure 6 shows that the "average synaptic weight of left thalamic sensory cells (TCR_L) to left cortical motor cells (PY_L) sections" and "average synaptic weight of right thalamic sensory cells (TCR_L) to right cortical motor cells (PY_R) sections" are less than "average synaptic weight of right thalamic sensory cells (TCR_R) to left cortical motor cells (PY_L) sections" and "average synaptic weight of left thalamic sensory cells (TCR_L) to right cortical motor cells (PY_R) sections". In other words, the pathways between sensors and motors on the same side are strengthened compared to those of the opposite side. It shows that the robot is successfully trained to approach the target. The distribution of synaptic weights in the trained network is shown in Fig. 7 (The maximum strength of synaptic weights is taken to be 3 mV). Figure 7 demonstrates that most of the synapses at the end of the training have weights less than 0.5 mV, while the weight of some synapses is around 3 mV, which are belonged to amplified pathways. Furthermore, Fig. 8 shows the spiking frequency of each subsection during the training procedure.

As it is shown in Fig. 8, the frequency of spikes of inhibitory spiking neurons is increased after 2×10^4 s of running time. This is because of the saturation of fast spiking inhibitory neurons regarding their physiological properties, but it does not affect the learning process. A sample diagram of STDP release which occurs after a spike and its duration is shown in Fig. 9.

As it is shown in Fig. 9, the effective STDP duration (the time in which variation of synaptic weight is possible) is about 100 ms, and after 80 ms, the amount of STDP will be very small to have effect on variation of synaptic weight (Eq. 6). Figure 10 shows the distance of the target from the

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

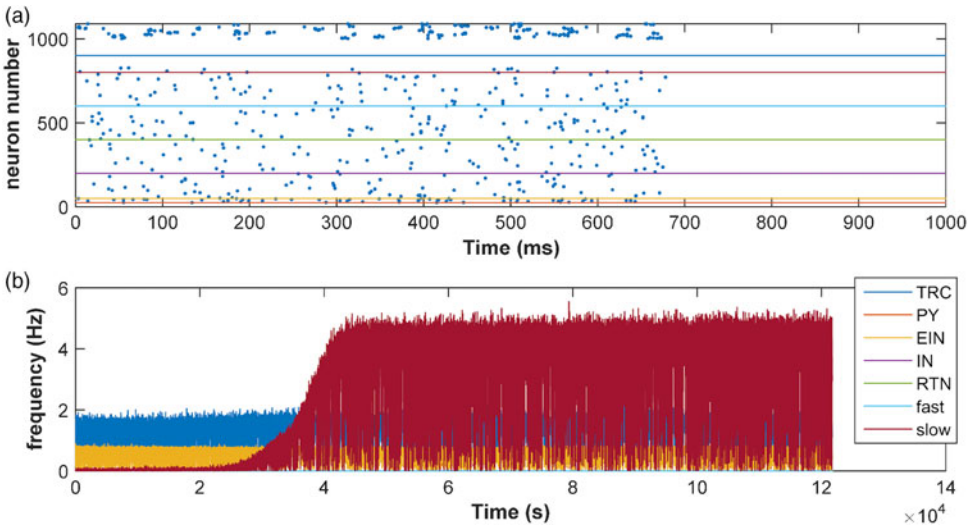

Fig. 8. (a) Sample of spikes of neural network, (b) Spiking frequency of subsections.

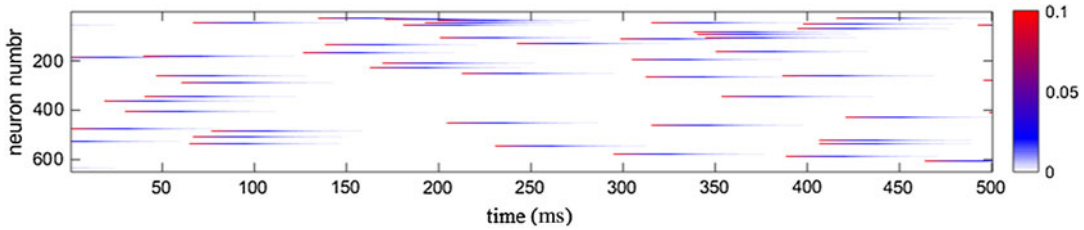

Fig. 9. Example of release and duration of STDP in synapses.

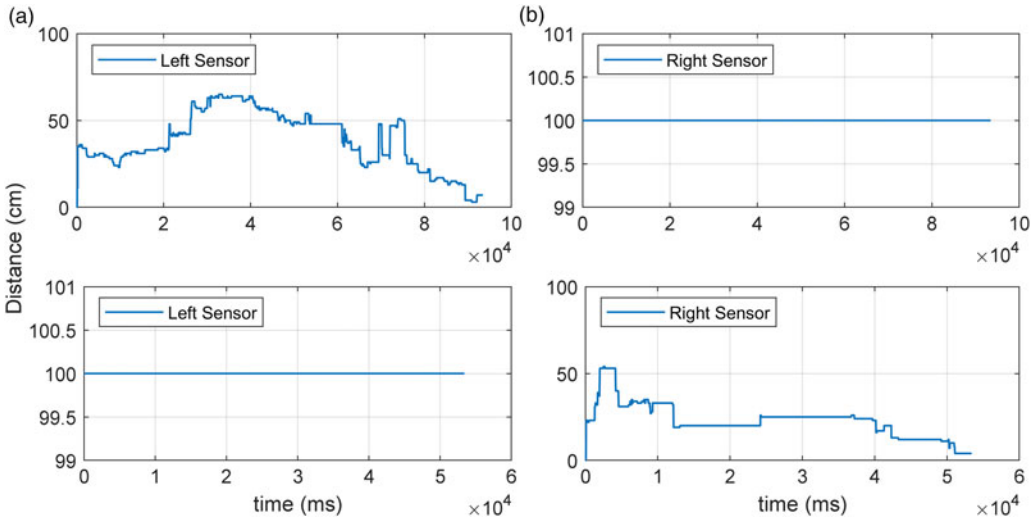

Fig. 10. Right and left sensor values in the experimental study. (a) Target is located on the left side of the robot. (b)Target is located on the right side of the robot.

robot in two case studies (after learning). In one case study, the target is located on the right side of the robot and in another one it is located on the left side.

As it is shown in Fig. 10, when the target is on the right side of the robot in an arbitrary distance of 30 cm, at first, the robot goes a little farther in the opposite side of the target. However, after some back and forth movements, the robot approaches the target and catches it. In this test, the left

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

Table V. Experimental study results.

Initial target distance from the robot	**20 cm**	**30 cm**	**40 cm**
Accomplishment time (target on the right side)	51±9 s	55±13 s	69±16 s
Accomplishment time (target on the left side)	89±12 s	81±16 s	92±19 s

sensor does not sense anything in the range of 100 cm, so the maximum value of 100 cm is depicted in the figure. Figure 10 shows the same state when the target is on the left side. This test proves that the robot has successfully learned the target attraction task. The reason that the measurements seem discontinues is that the robot moves discontinuously. Table V shows the test results for three different distances when the target is on the right or left side of the robot. In each test, the robot accomplished the target attraction task.

During experimental tests, it is observed that the overall latency in the system, including the communication latency and response time of the robot, changes the behavior of the system compared to the simulation scenario, where the program could run up to 4 kHz. Dividing movements of the robot into discontinuous periods solves this problem and effectively overcomes all the system delays (e.g., robot does not continue moving after executing one command and stops for a small time). Experimental studies show that the robot eventually converges to the target, even if it oscillates back and forth for a few seconds. These tests prove that through the proposed method, the thalamo-cortical structure is trained successfully to perform different tasks. This is valuable because the architecture assimilates the real architecture of the mammalian brain. Therefore, this ushers a very bright future for this kind of bio-inspired systems and proves that even more complex tasks will eventually be handled by this structure. However, more complex systems of mammalian brain, for example, basal ganglia and cerebellum, may be helpful for future studies. Here, two sensors and motors are considered for the robot, but it may be increased for detection of obstacle and more complex robot. In the case of mobile robots with rotation ability (e.g., when there are two differential wheels), the left and right sections in the "PY" area of neural networks are connected to the left and right motors of the robot. To evaluate the effectiveness of algorithm for differentially driven mobile robot, after learning process, the average synaptic weights of TCR_R section to PY_L section (and the average synaptic weights of TCR_L section to PY_R section) should be more than average synaptic weights of TCR_R section to PY_R section (and the average synaptic weights of TCR_L section to PY_L section).

5. Conclusion and Future Works

Neurons are basic blocks of information processing systems in nature. Connecting neurons results in complex networks that are used to process information in simpler organisms from C-elegance worm with 302 neurons to mammals with complex thalamo-cortical architecture. Recent studies in neuroscience show that circuits between thalamus and sensory–motor sections of cortex play an important role in coordinating motor learning,[59] and plasticity of thalamo-cortical projections onto neurons of sensory–motor cortex is effective in learned motor tasks.[60] Action learning of robots in an unknown environment is mostly related to "reinforcement learning" in robot–environment interaction and the findings of this paper, for the first time, studies it based on the TCT structure. A new reinforcement learning method based on the SNN model of the TCT architecture is investigated. The total number of 1090 neurons in two main sections of thalamus and cortex are considered in the network with an exact connection map. Spiking behavior of each neuron is modeled through the Izhikevich method. Dopamine-modulated STDP is used for learning algorithm in which dopamine delivery is related to external conditions, instead of internal neural pathways. The proposed architecture contains cortical module comprised of the left and right PY cells as motor neurons, TRCs as sensory cells, and inhibitory cell of cortex as TRN. The evidence from simulation studies alongside the experimental tests on a mobile robot named Tabrizbot points towards the feasibility of using more realistic models and architectures of SNNs in reinforcement learning of robots. The effectiveness of the proposed method is evaluated by comparing the results before and after the learning process in terms of "foraging time" and "variation of synaptic weights." After learning, the meantime of catching the target is decreased by about 36% and synaptic weights of specified neurons are increased according to the desired benchmark. This method could be developed for learning problems of fix robots, obstacle avoidance of mobile robots in cluttered environments, and concurrent target attraction-obstacle

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

avoidance tasks. Moreover, this method is very beneficial for the learning problem of large-scale robots, which have complex dynamics. In the future, we are planning to develop SNNs for applications in reinforcement learning of large-scale robots with more DOFs and more sensors. Also, we will study the integration of optimal control and reinforcement learning to teach path planning to a robot. Implementation of more complex algorithms, including the function of basal ganglia, is another interesting topic that is useful in the development of reinforcement learning systems for robots.

Acknowledgment

We have to express our appreciation to Dr. Basabdatta Sen Bhattacharya for assistance with TCT architecture and sharing their pearls of wisdom with us during this research.

References

1. S. Murray Sherman, "Thalamus," *Scholarpedia* **1**(9), 1583 (2006).
2. S. Grossberg and M. Versace, "Spikes, synchrony, and attentive learning by laminar thalamocortical circuits," *Brain Res.* **1218**, 278–312 (2008).
3. F. Chersi, M. Mirolli, G. Pezzulo and G. Baldassarre, "A spiking neuron model of the cortico-basal ganglia circuits for goal-directed and habitual action learning," *Neural Networks* **41**, 212–224 (2013).
4. M. Andrés Chalita, D. Lis and A. Caverzasi, "Reinforcement learning in a bio-connectionist model based in the thalamo-cortical neural circuit," *Biolog. Ins. Cogn. Arch.* **16**, 45–63 (2016).
5. T. C. Stewart, T. Bekolay and C. Eliasmith, "Learning to select actions with spiking neurons in the basal ganglia," *Front. Neurosci.* **6**, 2 (2012).
6. H. Shteingart and Y. Loewenstein, "Reinforcement learning and human behavior," *Curr. Opinion Neurobiol.* **25**, 93–98 (2014).
7. T. V. Maia, "Reinforcement learning, conditioning, and the brain: Successes and challenges," *Cogn. Affect. Behav. Neurosci.* **9**(4), 343–364 (2009).
8. B. W. Balleine, R. W. Morris and B. K. Leung, "Thalamocortical integration of instrumental learning and performance and their disintegration in addiction," *Brain Res.* **1628**(A), 104–116 (2015).
9. Y. H. Tanaka, Y. R. Tanaka, M. Kondo, S.-I. Terada, Y. Kawaguchi and M. Matsuzaki, "Thalamocortical axonal activity in motor cortex exhibits layer-specific dynamics during motor learning," *Neuron* **100**(1), 244–258 (2018).
10. E. M. Izhikevich, "Which model to use for cortical spiking neurons?," *IEEE Trans. Neural Networks* **15**(5), 1063–1070 (2004).
11. M. Breakspear, "Dynamic models of large-scale brain activity," *Nature Neurosci.* **20**(3), 340 (2017).
12. M. Sarim, T. Schultz, M. Kumar and R. Jha, "An Artificial Brain Mechanism to Develop a Learning Paradigm for Robot Navigation," *ASME 2016 Dynamic Systems and Control Conference* (American Society of Mechanical Engineers, 2016) pp. V001T03A004–V001T03A004.
13. E. M. Izhikevich and G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *Proc. Nat. Acad. Sci.* **105**(9), 3593–3598 (2008).
14. E. M. Izhikevich, "Solving the distal reward problem through linkage of STDP and dopamine signaling," *Cerebral Cortex* **17**(10), 2443–2452 (2007).
15. R. Elibol and N. S. Şengör, "Building neurocomputational models at different levels for basal ganglia circuit," *Istanbul Univ. J. Elect. Electron. Eng.* **17**(1), 3137–3146 (2017).
16. E. Erçelik and N. S. Şengör, "A Neurocomputational Model Implemented on Humanoid Robot for Learning Action Selection," *2015 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2015) pp. 1–6.
17. J. Kober, J. Andrew Bagnell and J. Peters, "Reinforcement learning in robotics: A survey," *Int. J. Robot. Res.* **32**(11), 1238–1274 (2013).
18. Z. Miljković, M. Mitić, M. Lazarević and B. Babić, "Neural network reinforcement learning for visual control of robot manipulators," *Expert Syst. Appl.* **40**(5), 1721–1736 (2013).
19. Z. Bing, C. Meschede, F. Röhrbein, K. Huang and A. C. Knoll, "A survey of robotics control based on learning-inspired spiking neural networks," *Front. Neurorobot.* **12**, 35 (2018).
20. M. Khamassi, S. Lallée, P. Enel, E. Procyk and P. F. Dominey "Robot cognitive control with a neurophysiologically inspired reinforcement learning model," *Front. Neurorobot.* **5**(1), 1–3 (2011).
21. T. J. Prescott, F. M. Montes González, K. Gurney, M. D. Humphries and P. Redgrave, "A robot model of the basal ganglia: Behavior and intrinsic processing," *Neural Networks* **19**(1), 31–61 (2006).
22. L. Long and G. Fang, "A Review of Biologically Plausible Neuron Models for Spiking Neural Networks," **In:** *AIAA Infotech@ Aerospace 2010* (2010) p. 3540.
23. M. Burrafato and L. Florio, "A cognitive architecture based on an amygdala thalamo cortical model for developing new goals and behaviors: application in humanoid robotics," Master's thesis (Politecnico di Milano, 2012).
24. B. S. Bhattacharya, C. Patterson, F. Galluppi, S. J. Durrant and S. Furber, "Engineering a thalamo-cortico-thalamic circuit on spinnaker: A preliminary study toward modeling sleep and wakefulness," *Front. Neural Circ.* **8**, 46 (2014).
25. B. S. Bhattacharya, D. Coyle and L. P. Maguire, "A thalamo–cortico–thalamic neural mass model to study alpha rhythms in alzheimers disease," *Neural Networks* **24**(6), 631–645 (2011).

26. J. Igarashi, O. Shouno, T. Fukai and H. Tsujino, "Real-time simulation of a spiking neural network model of the basal ganglia circuitry using general purpose computing on graphics processing units," *Neural Networks* **24**(9), 950–960 (2011).
27. M. D. Humphries, R. D. Stewart and K. N. Gurney, "A physiologically plausible model of action selection and oscillatory activity in the basal ganglia," *J. Neurosci.* **26**(50), 12921–12942 (2006).
28. K. Gurney, T. J. Prescott and P. Redgrave, "A computational model of action selection in the basal ganglia. i. a new functional anatomy," *Biolog. Cybern.* **84**(6), 401–410 (2001).
29. O. Shouno, J. Takeuchi and H. Tsujino, "A Spiking Neuron Model of the Basal Ganglia Circuitry that can Generate Behavioral Variability," **In:** *The Basal Ganglia IX* (H. J. Groenewegen, P. Voorn, H. W. Berendse, A. B. Mulder and A. R. Cools, eds.) (Springer, New York, 2009) pp. 191–200.
30. Z. Cao, L. Cheng, C. Zhou, N. Gu, X. Wang and M. Tan, "Spiking neural network-based target tracking control for autonomous mobile robots," *Neural Comput. Appl.* **26**(8), 1839–1847 (2015).
31. P. Arena, S. De Fiore, L. Patané, M. Pollino and C. Ventura, "Insect Inspired Unsupervised Learning for Tactic and Phobic Behavior Enhancement in a Hybrid Robot," *The 2010 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2010) pp. 1–8.
32. A. Bouganis and M. Shanahan, "Training a Spiking Neural Network to Control a 4-DOF Robotic Arm based on Spike Timing-Dependent Plasticity," *The 2010 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2010) pp. 1–8.
33. M. Nadjib Zennir, M. Benmohammed and R. Boudjadja, "Spike-Time Dependant Plasticity in a Spiking Neural Network for Robot Path Planning," *AIAI Workshops* (2015) pp. 2–13.
34. V. Azimirad, M. F. Sani and M. T. Ramezanlou, "Unsupervised Learning of Target Attraction for Robots Through Spike Timing Dependent Plasticity," *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)* (IEEE, 2017) pp. 0428–0433.
35. E. Nichols, L. J. McDaid and N. H. Siddique, "Case study on a self-organizing spiking neural network for robot navigation," *Int. J. Neural Syst.* **20**(06), 501–508 (2010).
36. R. Batllori, C. B. Laramee, W. Land and J. David Schaffer, "Evolving spiking neural networks for robot control," *Procedia Comput. Sci.* **6**, 329–334 (2011).
37. A. Cyr and M. Boukadoum, "Classical conditioning in different temporal constraints: An STDP learning rule for robots controlled by spiking neural networks," *Adapt. Behav.* **20**(4), 257–272 (2012).
38. X. Zhang, Z. Xu, C. Henriquez and S. Ferrari, "Spike-Based Indirect Training of a Spiking Neural Network-Controlled Virtual Insect," *52nd IEEE Conference on Decision and Control* (IEEE, 2013) pp. 6798–6805.
39. E. Nichols, L. J. McDaid and N. Siddique, "Biologically inspired SNN for robot control," *IEEE Trans. Cybern.* **43**(1), 115–128 (2012).
40. P. Mazumder, D. Hu, I. Ebong, X. Zhang, Z. Xu and S. Ferrari, "Digital implementation of a virtual insect trained by spike-timing dependent plasticity," *Integration* **100**(54), 109–117 (2016).
41. H. Masuta and N. Kubota, "Learnablity of a spiking neural network for perception of a partner robot," *2008 IEEE International Conference on Systems, Man and Cybernetics* (IEEE, 2008) pp. 1413–1418.
42. H. Hagras, A. Pounds-Cornish, M. Colley, V. Callaghan and G. Clarke, "Evolving Spiking Neural Network Controllers for Autonomous Robots," *IEEE International Conference on Robotics and Automation. ICRA'04.*, vol, 5 (IEEE, 2004) pp. 4620–4626.
43. F. Alnajjar and K. Murase, "A simple aplysia-like spiking neural network to generate adaptive behavior in autonomous robots," *Adaptive Behavior* **16**(5), 306–324 (2008).
44. N. Takase, J. Botzheim and N. Kubota, "Evolving Spiking Neural Network for Robot Locomotion Generation," *2015 IEEE Congress on Evolutionary Computation (CEC)* (IEEE, 2015) pp. 558–565.
45. Y. Oniz and O. Kaynak, "Control of a direct drive robot using fuzzy spiking neural networks with variable structure systems-based learning algorithm," *Neurocomputing* **149**(PB), 690–699 (2015).
46. X. Wang, Z.-G. Hou, A. Zou, M. Tan and L. Cheng, "A behavior controller based on spiking neural networks for mobile robots," *Neurocomputing* **71**(4–6), 655–666 (2008).
47. N. Singh, C. R. Huyck, V. Gandhi and A. Jones, "Neuron-based control mechanisms for a robotic arm and hand," *Int. J. Comput. Elect. Auto. Control Inf. Eng.* **11**(2), 221–229 (2017).
48. X. Wang, Z.-G. Hou, M. Tan, Y. Wang and X. Wang, "Corridor-Scene Classification for Mobile Robot Using Spiking Neurons," *2008 Fourth International Conference on Natural Computation*, vol. 4 (IEEE, 2008) pp. 125–129.
49. X. Wang, Z.-G. Hou, M. Tan, Y. Wang and L. Hu, "The Wall-Following Controller for the Mobile Robot Using Spiking Neurons," *2009 International Conference on Artificial Intelligence and Computational Intelligence*, vol. 1 (IEEE, 2009) pp. 194–199.
50. X. Wang, Z.-G. Hou, F. Lv, M. Tan and Y. Wang, "Mobile robots modular navigation controller using spiking neural networks," *Neurocomputing* **134**, 230–238 (2014).
51. L. I. Helgadottir, J. Haenicke, T. Landgraf, R. Rojas and M. P. Nawrot, "Conditioned Behavior in a Robot Controlled by a Spiking Neural Network," *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* (IEEE, 2013) pp. 891–894.
52. E. Dumesnil, P.-O. Beaulieu and M. Boukadoum, "Robotic Implementation of Classical and Operant Conditioning as a Single STDP Learning Process," *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2016) pp. 5241–5247.
53. S. Dura-Bernal, G. L. Chadderdon, S. A. Neymotin, J. T. Francis and W. W. Lytton, "Towards a real-time interface between a biomimetic model of sensorimotor cortex and a robotic arm," *Pattern Recogn. Lett.* **36**, 204–212 (2014).

54. E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Networks* **14**(6), 1569–1572 (2003).
55. W. Schultz, "Predictive reward signal of dopamine neurons," *J. Neurophysiology* **80**(1), 1–27 (1998).
56. P. Chorley and A. K. Seth, "Closing the Sensory-Motor Loop on Dopamine Signalled Reinforcement Learning," *International Conference on Simulation of Adaptive Behavior* (Springer, 2008) pp. 280–290.
57. S. A. Neymotin, G. L. Chadderdon, C. C. Kerr, J. T. Francis and W. W. Lytton, "Reinforcement learning of two-joint virtual arm reaching in a computer model of sensorimotor cortex," *Neural Comput.* **25**(12), 3263–3293 (2013).
58. M. Ursino, F. Cona and M. Zavaglia, "The generation of rhythms within a cortical region: Analysis of a neural mass model," *NeuroImage* **52**(3), 1080–1094 (2010).
59. A. J. Yonk and D. J. Margolis, "Traces of learning in thalamocortical circuits," *Neuron* **103**(2), 175–176 (2019).
60. Y. Takashima, M. Scanziani, J. M. Conner, J. S. Biane and M. H. Tuszynski, "Thalamocortical projections onto behaviorally relevant neurons exhibit plasticity during adult motor learning," *Neuron* **89**(6), 1173–1179 (2016).

Appendix

To achieve the highest efficiency, the program of the robot is implemented using RTOS and every different function is performed in its task. The simplified pseudocode of the robot's program is shown in Algorithm A1. Each task runs every 50 ms. Also, some other parts of the program are written in the interrupt-based mode to avoid blocking the rest of the program. For instance scanning, multiple ultrasonic sensors demand interrupt-based coding. In every sequence, the program triggers the ultrasonic sensor and starts the timer. When the emitted wave reflects from the obstacle and returns to the sensor, the program will read it from interrupt subroutine. furthermore, communication with the ESP module also requires many waiting sequences, and we overcome this problem by handling the related tasks in interrupt subroutines. The simplified flowchart of the robot is shown in Fig. A1.

Algorithm A1 TABRIZBOT simplified algorithm

```
1:  procedure READ SENSOR( )
2:      while 1 do
3:          right_sensor_value ← read_right_sensor()
4:          left_sensor_value ← read_left_sensor()
5:  procedure ESP COMMUNICATE( )
6:      while 1 do
7:          ESP_Send_To_PC(right_sensor_value, left_sensor_value)
8:          motor_speed ← ESP_Recieve_From_PC()
9:  procedure MOTOR DRIVE( )
10:     while 1 do
11:         if motor_speed > 0 then
12:             CWDrive_Left_Motor(motor_speed)          ▷ clock wise drive
13:             CWDrive_Right_Motor(motor_speed)
14:         else
15:             CCWDrive_Left_Motor(−motor_speed)        ▷ counter clock wise drive
16:             CCWDrive_Right_Motor(−motor_speed)
17: procedure LCD UPDATE( )
18:     while 1 do
19:         Lcd_show(ESP_status)
20:         Lcd_show(motor_speed)
21:         Lcd_show(right_sensor_value)
22:         Lcd_show(left_sensor_value)
```

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press

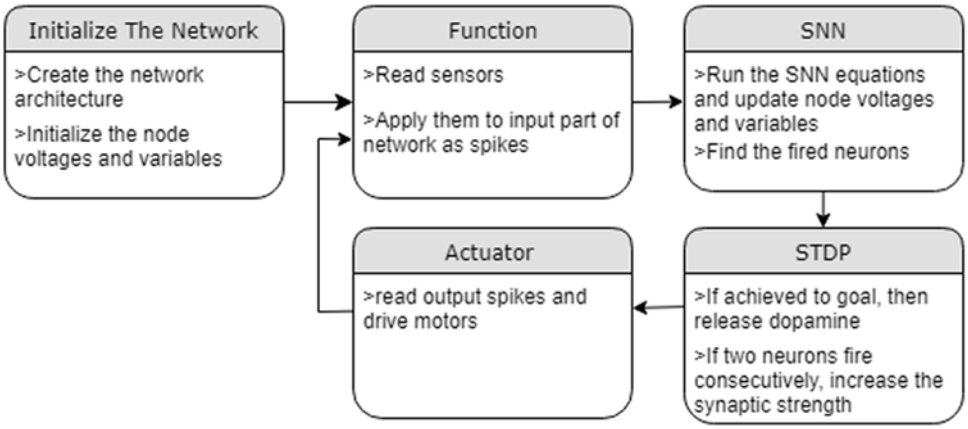

Fig. A1. Simplified flowchart of the implemented code for the experimental setup.

https://doi.org/10.1017/S0263574719001632 Published online by Cambridge University Press