Biologically Inspired Neural Controller for Robot Learning and Mapping

Alejandra Barrera Ramírez and Alfredo Weitzenfeld Ridel

***Abstract*—In this paper we present a model composed of layers of neurons designed on the basis of the neurophysiology of the rat hippocampus to control the navigation of a real robot. The model allows the robot to learn reward locations in different mazes and to return home autonomously by building a topological map of the environment. We described robotic experimentation results from our tests in a T-maze, an 8-arm radial maze and a 3-T shaped maze.**

I. Introduction

To explain the ability of rats to solve spatial problems, Tolman argued in 1948 that rats should have a **cognitive map** in some part of their brain [1]. Then, in 1978, O'Keefe and Nadel argued that such map was located in the **hippocampus** [2]. Experimental work has shown that there exist at least two distinct populations of neurons in the rat hippocampus known as **place cells** and **head-direction cells**. Place cells codify information about physical locations of the animal, while head-direction cells codify orientations of the animal's head [3]. Studies on the rat brain have provided inspiration in developing alternative robotic navigation models to those based on classical approaches, such as metric and topological [4]. Examples of such navigation models based on the rat's hippocampus neurophysiology are those by Burgess and O'Keefe [5], Touretzky and Redish [6], Balakrishnan, Bhatt and Honavar [7], Trullier and Meyer [8], Arleo and Gerstner [9], Gaussier, Revel, Banquet and Babeau [10], Guazzelli, Corbacho, Bota and Arbib [11], and more recently Milford and Wyeth [12].

In this paper we present a navigation model based on the neurophysiology of the rat hippocampus that allows an actual robot to learn reward locations in different mazes, to build a map based representation of the environment and to return home autonomously. Our model was tested in different learning and mapping experiments using motivation and reinforcement learning. Section 2 of the paper describes neurobiological experiments on spatial memory serving as the basis for our model, Section 3 presents the model description, Section 4 discusses the robotic architecture and experimentation results, and Section 5 presents conclusions.

This research was partially supported by collaboration projects UC MEXUS CONACYT (ITAM – UCSC), LAFMI CONACYT (ITAM – ISC), NSF CONACYT (ITAM – UCI) under grant #42440 and "Asociación Mexicana de Cultura, S. A."

Alejandra Barrera and Alfredo Weitzenfeld are with the Computer Engineering Department at the Instituto Tecnológico Autónomo de México. Río Hondo #1, Tizapán San Ángel, CP 01000, México DF, México (e-mail: abarrera@itam.mx, alfredo@itam.mx).

II. Neurobiological Experiments

Among the neurobiological experiments that have been done with rats solving spatial tasks, such as [13] and [14], O'Keefe [15] demonstrated that rats with damage to the hippocampal system can still learn this kind of tasks arguing that, besides the spatial system located within the hippocampus, there are other related ones located outside the hippocampus. Furthermore, O'Keefe argued that many spatial tasks can be learned using more than one of them, such as the egocentric orientation system that specifies behavior in terms of rotations within a given framework centered on the body midline (see Fig. 2 left).

In order to explore the properties of the egocentric orientation system, O'Keefe experimented with the reversal task in a T-maze and in an 8-arm radial maze. The experiment consisted in training rats to turn to the left arm of the "T" by rewarding them at the end of that arm. When rats learned the correct turn, the reward was moved to the end of the opposite arm. As a result, rats had to unlearn the previous arm to learn the new one. During the reversal task an 8-arm probe was introduced every third trial, enabling O'Keefe to evaluate the rats' orientation during relearning. For lesioned animals the results indicated that in the T-maze there was an abrupt shift between incorrect and correct reward arm, but in the 8-arm radial maze the shift in the rats' orientation was incremental from the starting quadrant (-90°, -45°) through straight ahead (0°) and into the new reversal quadrant (+45°, +90°). The conclusion about lesioned rats' behavior was that the choice in the T-maze was the manifestation of the underlying orientation vector.

On the other hand, the performance of normal rats in the T-maze proceeded in the same way as in lesioned rats, but the underlying orientation vector of normal rats did not shift in a smooth manner but jumped around randomly. O'Keefe concluded that the reversal performance of normal rats was not based only on their orientation system, but also on the use of the hippocampal cognitive mapping system. Fig. 9 shows the average performance of four normal rats during the reversal phase of the experiment in the 8-arm radial maze. As can be seen, from the starting reward arm (-90°) and in early trials of the reversal phase rats picked different arms randomly until they begin to choose arms located around the new reward arm (+90°).

III. Model Description

The model we have developed controls the navigation of the rat by determining the direction of its movement and by building a map-based representation of the environment. The model is composed of layers of neurons that implement Hebbian [16] and reinforcement learning [17] in order to allow the expression of goal-oriented behavior. Fig. 1 shows the different layers in the model. The sensory inputs to the model are composed of: (i) **affordances**, (ii) **kinesthetic information** and (iii) the **internal drive** of the rat.

The main layers of the model are the **place cell layer** (PCL) and the **world graph layer** (WGL). Place cell activity is influenced by path integration information. The pattern of kinesthetic information generated by **path integration feature detector layer** (PIFDL) is the input to PCL. The pattern of activity generated in this layer represents a single place or location in the environment.

The world graph is implemented in the model by WGL, whose nodes are created on demand. Every neuron in PCL is connected to every node in WGL. Each WGL node can store eight different activity patterns, one for each direction, assuming that the animal can orient itself in eight directions (see Fig. 2 right) and can experiment different views from the same place. The arcs in the graph are represented by links between two WGL nodes, and are associated with the orientation of the rat's head when the animal goes from one node to the next one. In our model, the determination of the direction of the rat's head is based on a global coordinate system, which is relative to an environmental anchor that corresponds to the departure location in the exploration process. Fig. 2 (right) shows this global coordinate system.

The model is able to associate expectations of future reward to specific locations in the environment in order to allow the rat to learn goal locations. To do this, the model incorporates a **motivational schema** and an **action selection schema**, which determines the next direction of the rat's head. We discuss the model components, the animal's motivation and learning in the following subsections.

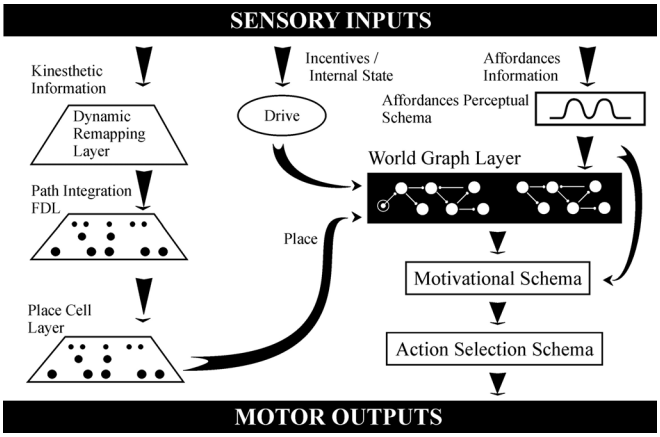

Fig. 1. The layers of the hippocampus-based navigation model. FDL stands for Feature Detector Layer.

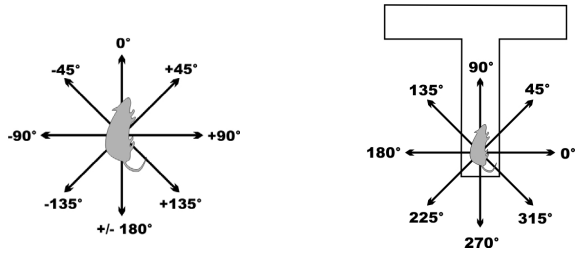

Fig. 2. (Left) Local coordinate system used in the model to determined the relative affordances for movement. (Right) Global coordinate system used in WGL to build the map-based representation of the external environment; e.g., a T-maze. This global system is relative to the departure location (the base of the T in the figure).

A. Affordances

The term **affordance**, adopted from Gibson [18], refers to sensory information that an animal uses to interact with the environment without the need to recognize objects. Specifically, the notion of **affordances for movement** represents all possible motor actions that a rat can execute through the immediate sensing of its environment; e.g., visual sighting of a corridor – go straight ahead, sensed branches in a maze – turn, etc.

In our model, affordances for movement are coded in a linear array of cells called an **affordances perceptual schema** (APS) that represents possible turns from -180° to +180° in 45° intervals. In this way, when the rat is in the center of an 8-arm radial maze, it is able to visually sense eight different arms and consequently perceive eight different affordances (nine if we consider that -180° and +180° are represented separately). Determination of the affordances for movement is based on a local coordinate system that is relative to the rat's head, as shown in Fig. 2 (left).

APS is represented as an array of neurons whose activation level is computed through en exponential equation. For a specific affordance (*aff*), the activation level of neuron *i* is determined as follows:

$$aff_i = e^{-\frac{(i-a)^2}{2d^2}}, \quad (1)$$

where *d* is a constant value representing the width of the exponential, and *a* is a constant that depends on the relative direction of the affordance: from -180° to +180°, $a = 4+9m$ with *m* from 0 to 8.

Every available affordance has the form shown in (1); e.g. the information picked up by the APS when the rat is in the center of an 8-arm radial maze is a sum of all available affordances as shown in Fig. 3 and the value of each of its neurons is computed using (2).

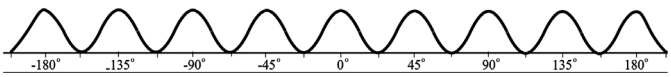

Fig. 3. Affordances perceptual schema when the rat is in the center of an 8-arm radial maze.

$$aff_i = \sum_{m=0}^{8} e^{-\frac{(i-(4+9m))^2}{2d^2}} \quad (2)$$

B. Kinesthetic Information

The second kind of sensory input to the model is determined by kinesthetic information; i.e. rat internal body signals generated during rat's locomotion. These signals are used to carry out the path integration process that influences the place cells' activity, where **path integration** describes the process by which kinesthetic information allows the animal to update the position of its point of departure (an environmental anchor) each time it moves in relation to its current position. In this way, path integration allows the animal to return home. As can be seen in Fig. 1, the model includes a **path integration module** composed of a **dynamic remapping layer** (DRL) defined as a two-dimensional perceptual schema representing the particular environment and anchor coordinates, and a **path integration feature detector layer** (PIFDL) where the activation of its neurons constitutes a pattern of kinesthetic information.

DRL generates a dynamic remapping perceptual schema (DRPS) defined as a two-dimensional array of neurons. Initially, this perceptual schema codifies the position of the departure location using the following equation:

$$dr_{i,j} = e^{\frac{-(i-y)^2}{2d^2}+\frac{-(j-x)^2}{2d^2}}, \quad (3)$$

where *i, j* identifies the neuron within the two-dimensional array; *x*, *y* are the codification in the DRPS of the rat's coordinates in the environment.

The perceptual schema is updated each time the rat moves by displacing the anchor position in the same magnitude but in opposite direction to the rat's movement. Every neuron in DRPS is randomly connected to 50% of the neurons in PIFDL. Connection weights between layers are randomly initialized and normalized between 0 and 1 according to

$$w_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}}, \quad (4)$$

where w_{ij} is the connection weight from neuron *i* in DRPS to neuron *j* in PIFDL, and *n* is the total number of neurons in DRPS connected to neuron *j*.

The activation level A_j of neuron *j* in PIFDL is computed by adding the products between each input value I_i coming form neuron *i* in DRPS and the corresponding connection weight as follows:

$$A_j = \sum_{i=1}^{n} I_i w_{ij} \quad (5)$$

We use Hebbian learning to update the connection weights between layers, modeling the association between the activation of a group of neurons in DRPS, the activation of a specific set of neurons in PIFDL and the inhibition of others as follows:

$$\Delta w_{ij} = \alpha I_i w_{ij} G_j \quad (6)$$

where Δw_{ij} is the change in the weight w_{ij}, α is the learning rate, I_i is the input value coming from neuron *i* in DRPS, and G_j is a new activation value of neuron *j* that depends on the number of the place occupied by the original activation level A_j within the hierarchy of activation levels in PIFDL. Weights are normalized as follows:

$$w_{ij} = \frac{w_{ij} + \Delta w_{ij}}{\sum_{i=1}^{n} (w_{ij} + \Delta w_{ij})}, \quad (7)$$

When a group of connection weights into a PIFDL unit is increased, the rest of the connection weights to the same neuron are decreased allowing the development of groups of PIFDL units that respond to specific situations.

C. The Internal Drive

The animal's motivation is related to its internal need to eat, which is represented in the model as the hunger drive. In general, drives *d* can be appetitive or aversive. The idea is that each appetitive drive spontaneously increases with every time step towards $d_{\max}$, while aversive drives are reduced towards 0, both according to a factor α_d intrinsic to the animal. An additional increase occurs if an incentive *I(d,x,t)* is present such as the sight of food. Drive reduction *a(d,x,t)* takes place after food ingestion. If the animal is at place *x* at time step *t*, and the value of drive *d* at that time step *t* is *d(t)*, then the value of *d* for the animal at time step *t+1* will be

$$d(t+1) = d(t) + \alpha_d \left| d_{\max} - d(t) \right| - a(d,x,t) \left| d(t) \right| + I(d,x,t) \left| d_{\max} - d(t) \right| \quad (8)$$

The amount of reward the animal gets by the presence of food is dependent on its current motivational state. If the rat is extremely hungry, the presence of food might be very rewarding, but if not, it will be less rewarding. In this way, the reward value depends on the current value of the animal's drive *d(t)* and its corresponding maximum value, according to:

$$r(t) = \frac{d(t)}{d_{\max}} \quad (9)$$

D. Motivation and Learning

Schultz et al. [19] hypothesized that dopamine neurons in the brain may mediate the role of unexpected rewards to bring about learning, whereas neurons in the striatum and in structures projecting to the striatum respond to primary rewards in well-established behavioral tasks to maintain rather than bring about learning. The input layer of the striatum is divided into regions called striosomes surrounded by matrix regions. In 1995 Houk et al. [20] proposed that the striatum implements an actor-critic architecture [21], [17]. In such a system, a striosomal module would serve as a critic (an adaptive critic), while the matrix modules would serve as actors. In a reinforcement learning system implemented by an actor-critic architecture, primary rewards may occur only after a sequence of correct actions has ended, which may take a long time. Because of this, an actor-critic architecture

processes expected values of future reinforcements and the output of an adaptive critic unit is a prediction of the value of future reinforcement. In our model this predicted value is used to generate a reinforcement signal that is transmitted to WGL in order to reinforce eight actors associated to every node in the world map.

The adaptive critic unit receives the activity pattern registered in PCL as input and computes the prediction value of future reinforcement using

$$p(t) = \sum_{i=1}^{n} v_i \ , \tag{10}$$

where *p(t)* is the predicted reinforcement value at time *t*, v_i is the connection weight between neuron *i* in PCL and the adaptive critic unit. The weights added correspond to the most active neurons in PCL (we consider n= 5 neurons).

The connection weights to the critic unit are initialized to 0 and updated according to

$$v_i(t+1) = v_i(t) + \beta\, \hat{r}(t)\, \bar{x}_i \ , \tag{11}$$

where β is the learning rate; $\bar{x}_i$ is the elegibility trace of the activity level of each neuron in PCL; and $\hat{r}(t)$ is the temporal difference error between any two adjacent predictions and is computed using

$$\hat{r}(t) = r(t) + \gamma\, p(t) - p(t-1) \ , \tag{12}$$

where *r(t)* is the reward value that corresponds to the drive value at time *t*, and γ is a constant.

In every model iteration the reinforcement learning process starts by updating $\bar{x}_i$ values associated to the most active neurons in PCL. If the rat perceives food from its location, the elegibility trace is increased, otherwise is decreased.

As previously mentioned, the actor modules are associated to the map nodes and in particular to the eight different directions the rat can orient to. The eight actors represent the expectations of finding reinforcement in orienting to a certain direction at the current location. Every actor is implemented in a map node as a pair of "weight – elegibility trace". The reinforcement process is carried out in the actors when the reinforcement in the adaptive critic unit has finished, and consists in increasing the elegiblity trace associated to the current rat's head direction in the active node in the map. Then, the actor weights are updated for all map nodes using:

$$w_k^d(t+1) = w_k^d(t) + \alpha \hat{r}(t) e_k^d(t) \ , \tag{13}$$

where $w_k^d(t+1)$ is the actor weight associated to node *k* and direction *d* at time *t+1*, α is the learning rate, $\hat{r}(t)$ is the reinforcement signal generated by the critic unit at time *t*, and $e_k^d(t)$ is the elegibility trace of actor *d* in node *k* at time *t*.

At a given location, the choice of the rat to turn to a specific direction is influenced by a general rewarding signal generated through the expectations of future reward w_k^d of the actors. This rewarding signal is computed by the motivational schema in the model considering information that proceeds from the WGL. In this layer three nodes in the map are considered: the active one and two more in sequence. For each node the expectations of reinforcement values associated to the directions of the arcs pointing to other nodes are reviewed and the direction with the highest expectation value is selected. The different directions selected over the sequence of nodes as well as the corresponding expectation of reinforcement values are stored and sent to the motivational schema.

The mission of the motivational schema is to compute the input to the action selection schema, which is composed of four signals represented as perceptual schemas: the affordances perceptual schema (*aff*), the food perceptual schema (*tf*), the global expectation of future reinforcement perceptual schema (*ger*), and the curiosity perceptual schema (*cl*).

The food perceptual schema codifies the rotation the animal has to carry out to orient to food when it is visible from current location. When food is not visible, the rotation magnitude is determined randomly between available affordances. In this way, a random component is induced to the process of choosing a direction to orient to.

Through *cl* the model considers the fact that the animal may go to places that are not yet represented in the world map. In this way, if the rat is not motivated to go towards a previously experienced place, it will tend to choose, based on its curiosity level, an affordance that leads to a place not yet represented in the map. In this perceptual schema an exponential term like the one shown in (1) is added, corresponding to each available affordance associated to the direction of an arc not represented in the active node in the map.

The motivational schema uses the expectations of reinforcement values and the corresponding directions selected by WGL over the sequence of nodes to generate an expectation of future reward perceptual schema *efr*. The rotation the rat has to execute to orient to each direction is computed and added as an exponential term in this perceptual schema with strength depending on the expectation of reinforcement value associated with the direction, as shown in (14).

$$efr_i = v_j\, e^{-\frac{(i-a)^2}{2d^2}} \ , \tag{14}$$

where efr_i is the activity level of neuron *i* in *efr*, v_j is the expectation of reinforcement value corresponding to direction *j* (the height of the exponential), *d* is a constant value representing the width of the exponential, *a* is a constant that depends on the relative rotation: from -180° to +180°, *a*= 4+9*m* with *m* from 0 to 8.

There can be at most three exponential terms associated to the three nodes in the sequence. In order to generate a global

expectation of reinforcement signal (*ger*) that will influence the next behavior of the rat, the center of mass is computed. If there is no available affordance coinciding with the center of mass, it is moved to the neuron that corresponds to the selected direction from the active map node. The four perceptual schemas are added to generate the input to the action selection schema:

$$I_i = aff_i + tf_i + cl_i + ger_i \quad (15)$$

The action selection schema determines the next direction of the rat's head, from 0° to 315°, by considering the highest activation value in *I* and the value of the current direction of the rat. This schema also computes the angle by which the rat has to turn to point to the next direction, and the displacement the rat has to undergo to reach its next position. If the next direction of the rat is different from the current one, the rat will not move and the displacement is set to 0 giving the rat the opportunity to perceive a different view from the same location. Finally, when the rat is returning to the departure point after having finished a trial in the experiment, the action selection schema computes the next rat's direction form the built world map, reading the directions of the arcs that link the sequence of nodes from the departure location to the place where the returning process begins. In a previous paper we documented in detail the return process carried out in the model [22].

IV. Robotic Architecture and Experimentation Results

We tested the model using a Sony AIBO ERS-210 4-legged robot having a local camera. The model was designed and implemented using the **NSL system** [23] and can interact with a virtual or real environment through a **visual processing** module that takes as input the image perceived by either a virtual or real robot and a **motor control** module that executes rotations and/or translations on the virtual or real robot. We have used three different virtual and physically built experimental environments: a T-maze, an 8-arm radial maze and an extended maze. Corridors were built using a very resistant plastic material with 100 cm of length, and 35 cm of width and height. Different color papers were pasted over the walls of the mazes to simplify junction, food and end of corridors recognition.

The experiment carried out in the T-maze and in the 8-arm radial maze is inspired on the reversal task implemented by O'Keefe and described in Section 2. Our main goal was not the replication of this experiment but to extend it into more complex robotic mazes. Besides, we did not pursue to model the differences between normal and hippocampus-lesioned rats, but to develop a robotic model inspired on the neurophysiology of normal rat's brain.

As we described in Section 2, O'Keefe divided the experiment of the reversal task in two phases: training and testing. The training phase was carried out in the T-maze and the testing phase was carried out combining trials in the T-maze with trials in the radial maze. We decided for simplicity to implement the reversal task in both the T-maze and the radial maze separately. We have also implemented the reversal task within an environment where the reward is not visible at the first choice point of the maze, but at the second one. The following sections describe the robotic experimentation results obtained in the three mazes.

A. Experiment I: T-Maze

In the T-maze the rat navigates from the base of the T to either one of the two arm extremes, and then it returns to the departure location autonomously. This process is repeated in every experiment's trial. At each step in the experiment, the rat takes three pictures of the environment: the first one in the current head direction, the second one 90º to the right and the third one 90º to the left.

During the experiment, the rat builds the world graph map shown in Fig. 4. It is composed of seven nodes, each one created when the rat sensed a change in the available affordances and did not recognize the information pattern generated in the PCL of the model. In this way, the base of the T-maze is represented in the map by three nodes. The number 1 node corresponds to the place of departure "a;" node 2 represents the locations "b," "c" and "d," and node 3 corresponds to the location "e," the junction of the "T", where the rat decides to turn left or right. Each arm of the T-maze is represented by two nodes. The far most node in each arm (nodes 5 and 7) corresponds to the end of the corridor (left location "h" and right location "k"), while the previous one (nodes 4 and 6) corresponds to locations between the junction and the end left locations "g" and "f" and right locations "I" and "j").

During the training phase, the food is placed in the left arm of the maze. When the rat reaches the T junction, the sight of food makes it decide to turn left. The rat repeats this process for some trials (see Table I).

When the testing phase begins, the food is moved to the right arm. Since the rat is not motivated to turn left anymore, its curiosity level for the right arm, not yet represented in the map, and the sight of food at the end of that corridor make the rat explore it. In the following trials, the rat goes through an unlearning process, where the expectations of future reward in the model for the left arm will decrease continuously. During this process, the rat turns to the left during every trial. Finally, in trail 20 of the experiment the rat decides to turn right, starting a relearning process. In the beginning of this process, the expectations of future reward for the right arm are smaller than the combined curiosity and noise levels thus the rat tends to choose the left or right arm randomly. From trail 32 the expectations of reward for the right arm are the dominant influence in the behavior of the rat, making it choose the right corridor consistently.

After finishing a trial, the rat returns home by reading the directions stored in the map [22]. Fig. 5 shows pictures of the robot's behavior during the experiment and a "shortened" video can be found in our web site [24].

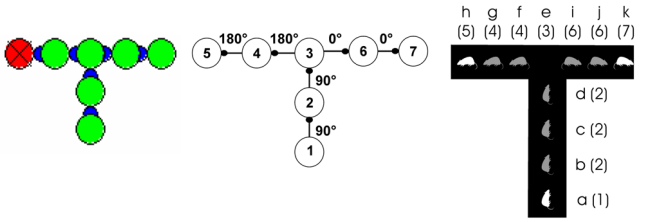

Fig. 4. The world map built by the rat during the reversal task in the T-maze. Nodes are numbered in order of creation. Arcs between nodes show the orientation of the rat when it moved from one node to the next one. In the T-maze shown, the different locations are labeled with letters and associated to the nodes of the map.

TABLE I. THE PERFORMANCE OF THE RAT DURING THE T-MAZE EXPERIMENT.

Trial #	Chosen arm	Phase	Process
1 – 10	Left	Training	Learning
11	Right	Testing	Curiosity drive
12 – 19	Left	Testing	Unlearning
20 – 31	Left or right randomly	Testing	Relearning
32 – …	Right	Testing	Relearning

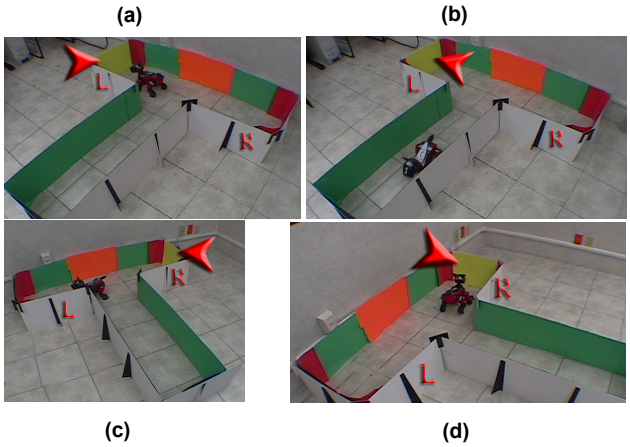

Fig. 5. (a) A typical trial in the training phase: the food is at left (L) and the robot is approaching it. (b) The robot is returning from the left arm. (c) A trial during the unlearning process: the food is at right (R) and the robot is approaching the end of the opposite corridor (L). (d) A trial during the relearning process: the food is at right and the robot is approaching it.

B. Experiment II: 8-Arm Radial Maze

In the 8-arm radial maze the rat navigates from the 270° arm to any other arm extreme, and then it returns to the departure location autonomously. This process is repeated in every experiment's trial. Fig. 6 presents the world graph map built by a rat during the experiment. Every arm is represented in the map by two nodes, and the choice point, by one node.

We divided the experiment in three phases: training, pre-testing and testing. The training phase works as in the T-maze with food placed at 180° arm (see Table II). During the pre-testing phase, the food is removed from the maze, and the rat explores other arms not represented in the map based on its curiosity drive; then, the rat goes through an unlearning process. In the testing phase, the food is placed at the end of the 0° arm and the relearning process begins making the rat to choose the 180° or 0° arm randomly. From trail 18 the expectations of reward for the 0° arm are the dominant influence in the behavior of the rat, making it choose this corridor consistently.

Fig. 7 shows pictures of the robot's behavior and a video can be found in [24].

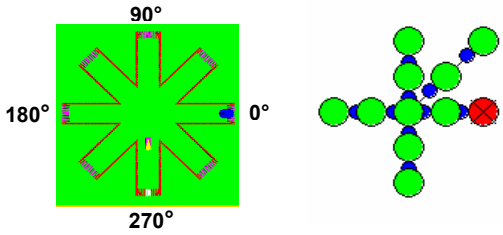

Fig. 6. The virtual environment used to test the reversal task in an 8-arm radial maze and the world map built by the rat.

TABLE II. THE PERFORMANCE OF THE RAT IN AN 8-ARM RADIAL MAZE.

Trial #	Chosen arm	Phase	Process
1 – 5	180°	Training	Learning
6 – 10	315°, 45°, 90°, 225° or 0° randomly	Pre-testing	Randomness and curiosity drives
11 – 12	180°	Pre-testing	Unlearning
13 – 17	180° or 0°	Testing	Relearning
18 – 22	0°	Testing	Relearning

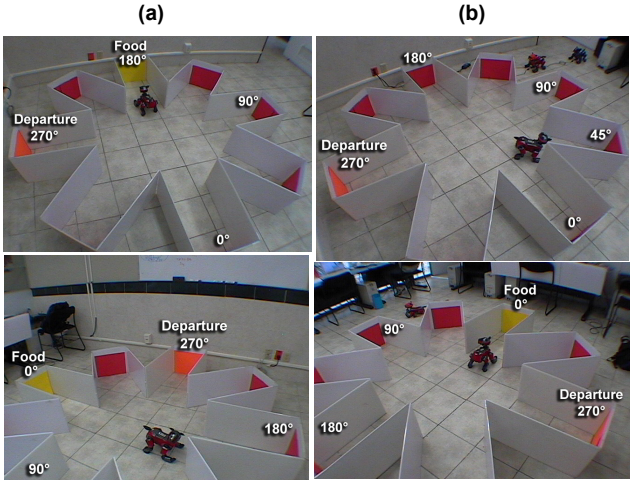

Fig. 7. (a) A typical trial in the training phase: the food is at 180° arm and the robot is approaching it. (b) A trial in the pre-testing phase: the robot chooses any arm not yet visited. (c) A trial in the testing phase: the food is at 0° arm and the robot is approaching the end of the previous reward arm. (d) The robot is approaching the new reward arm (0°).

In terms of learning, we can say that our results match qualitatively with those obtained by O'Keefe experimenting with normal rats despite some variations done to the original experiment. As in O'Keefe's rats, the underlying orientation vector of our rats did not shift in a smooth manner but jumped around randomly as shown in Fig. 8. Note that this graph is similar to the one displayed in Fig. 9.

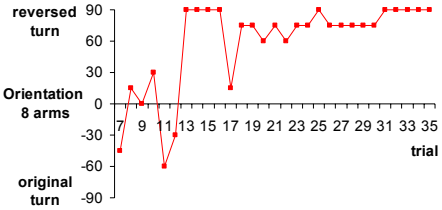

Fig. 8. Average performance of four simulated rats during the pre-testing and testing phase of the experiment in the 8-arm radial maze reported by our navigation model. Rats turn randomly to different arms between the original turn and the reversed turn.

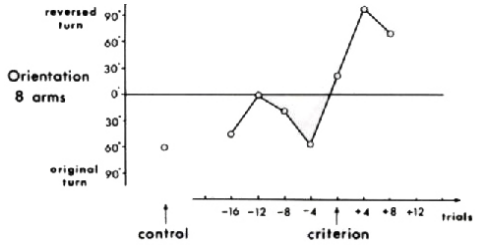

Fig. 9. Average performance of four normal rats during the reversal phase of the experiment in the 8-arm radial maze reported by O'Keefe. The graph is taken from [15]. Rats turn randomly to different arms between the original turn and the reversed turn.

C. *Experiment III: Extended Maze*

After completing the reversal task using a T-maze and an 8-arm radial maze, we decided to extend the experiment by considering a maze where the food was not visible by the rat at the first choice point but at the second one. To try this, we designed a maze composed of two horizontal Ts based on the arms of one vertical T. Fig. 10(a) presents the virtual environment used to simulate the model in this maze. During the experiment the rat navigates from the base of the vertical T to either one of its two arm extremes (0° or 180°) and then to either one of the two arm extremes of the corresponding horizontal T (90° or 270°). Then the rat returns to the departure location autonomously. This process is repeated in every experiment's trial. There are a training phase and a testing phase. In the first one food is placed at the end of the right arm (90°) of the left horizontal T (180°), and in the second one, the food is moved to the end of the right arm (270°) of the right horizontal T (0°).

Since the rat does not see food at the first choice point, it will explore the left or right arm randomly. In the training phase, only if the rat chooses the left arm, it would reach the second choice point from where food would be perceptible and the model would reinforce the right turn positively. However, in the next trial, there would not be a positive reinforcement for the left arm at the first choice point, thus the rat would pick the left or right arm randomly. Therefore, the rat would eventually learn to turn right at the second choice point, but could not learn to turn left at the first one. To solve this situation, we extended the model to reinforce reward paths instead of reward places. To do this, while the rat returns from a trial, the nodes connected in the map built from the goal back towards the departure place are positively reinforce by increasing in each map node the elegibility trace for the actor corresponding to its arc direction. If the goal is not reached by the end of a trial, the path is negatively reinforced. Considering the map built during the experiment shown in Fig. 10(b, c), and supposing that the rat has reached the goal place in the training phase, the reinforcement process consists on increasing the following elegibility traces in sequence: actor 90° in node 6, actor 90° in node 5, actor 180° in node 4, actor 180° in node 3, actor 90° in node 2 and actor 90° in node 1.

Table III summarizes the performance of the rat during the experiment. When the rat has chosen the left arm at the first choice point for at least 5 trials, the expectations of future reward for that arm are high enough to make the rat turn left in every training trial. In the first testing trial the rat turns left (1st choice point) – left (2nd choice point) motivated by curiosity. Then the rat goes through an unlearning and relearning process, and finally, in trial 24 the rat turns consistently right – right to get reward. Fig. 11 shows pictures of the robot's behavior and a video is found in [24].

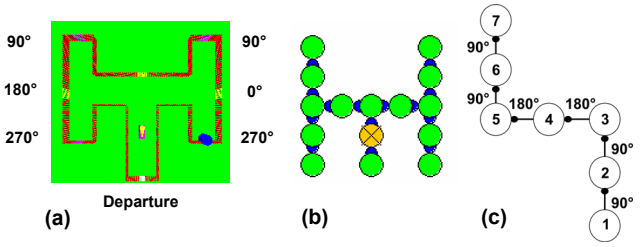

Fig. 10. (a) The virtual environment used to test the model in a 3-T shaped maze. (b) The world map built by a simulated rat. (c) The actors whose elegibility traces were reinforced positively during the training phase.

TABLE III. THE PERFORMANCE OF THE RAT IN A 3-T SHAPED MAZE.

Trial #	Chosen arm (1st choice point)	Chosen arm (2nd choice point)	Phase	Process
1 – 15	Left (180°) or right (0°) randomly	If the 1st choice was right (0°), the 2nd one is left (90°) or right (270°) randomly. If the 1st choice was left (180°), the 2nd one is right (90°).	Training	Positive path reinforcement during at least five trials left (180°)–right (90°)
16	Left (180°)	Left (270°)	Testing	Curiosity drive
17 – 18	Left (180°)	Right (90°)	Testing	Unlearning
19 – 23	Left (180°) – Right (90°) or Right (0°) – Right (270°)		Testing	Relearning
24 – 32	Right (0°)	Right (270°)	Testing	Relearning

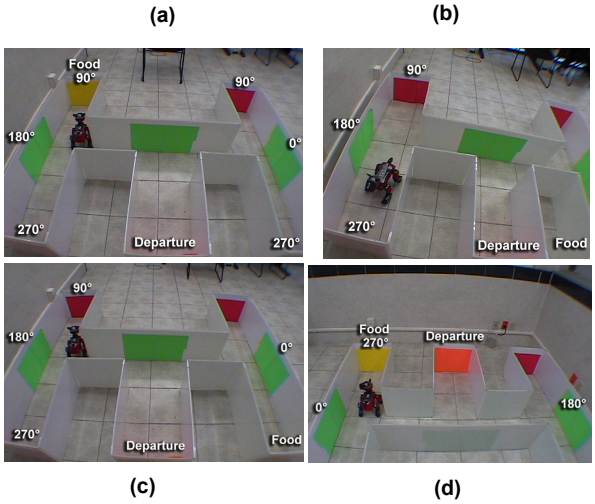

Fig. 11. (a) A trial in the training phase: the robot is approaching the food in the 180° arm. (b) In the testing phase, the robot explores a corridor not yet visited. (c) A trial during the unlearning process: the food is at 0° arm, but the robot still searches for it at the previous reward arm. (d) A trial during the relearning process: the robot is approaching the new reward location.

V. Conclusions

In this paper we have presented a robotic navigation model based on the physiology of the rat's brain. We have described the theoretical model followed by robotic experimentation results. We have shown that the rat/robot is able to explore a T-maze, an 8-arm radial maze and a 3-T shaped maze motivated by its internal need to eat, learn the locations of food, build a map of the environment and read the map to return to its departure location autonomously. The model achieves those goals using Hebbian and reinforcement learning.

We designed and tested the navigation model considering a real experiment carried out by O'Keefe in 1983 [15] and, in terms of learning, we can conclude that our results match qualitatively with those obtained from experimenting with normal rats. As part of this work, we also extended the reinforcement learning by reinforcing reward paths and not only reward places in mazes where reward is not visible by the rat from the first choice point. In this case we experimented with the 3-T shaped maze showing good results.

In our experiments the rat had only one way to reach the goal location because the departure location was the same in every trial. However, if we consider that the rat can reach the goal location from two different departure points in different trials of the experiment, then the reading of the map will need to be extended in enabling the rat to return home. Consequently, at this point we are considering to extend the model to use the path integration module to implement the return home process. Some other extensions that we plan to implement include the use of spatial landmarks for guiding rat navigation, and the adaptation of the model to solve cyclical mazes.

Finally, it should be emphasized that the motivation behind this work is the quest for inspiration from animal neurophysiology in solving spatial problems, as in the case of rats, where they have shown advanced learning capabilities that we expect will enhance robotic navigation models.

References

[1] E. Tolman. "Cognitive maps in rats and men." Psychological Review 55, pp. 189-208, 1948.

[2] J. O'Keefe and L. Nadel. "The hippocampus as a cognitive map." Oxford University Press, 1978.

[3] C. Hölscher. "Time, space and hippocampal functions." Reviews in the Neurosciences, 2003.

[4] A. Barrera and A. Weitzenfeld. "Biologically-inspired robotic mapping as an alternative to metric and topological approaches." Proceedings of the 1st IEEE Latin American Robotics Symposium, pp. 82-87. Mexico City, Mexico. October 28 – 29, 2004.

[5] N. Burgess, M. Recce and J. O'Keefe. "A model of hippocampal function." Neural Networks, Vol. 7, Nos. 6/7, pp. 1065-1081, 1994.

[6] D. Touretzky and A. Redish. "A theory of rodent navigation based on interacting representations of space." Hippocampus 6, pp. 247-270, 1996.

[7] K. Balakrishnan, R. Bhatt and V. Honavar. "A computational model of rodent spatial learning and some behavioral experiments." In: M. A. Gernsbacher & Sharon J. Derry (Eds.). Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Assoc., 1998.

[8] O. Trullier and J-A. Meyer. "Animat navigation using a cognitive graph." Biological Cybernetics 83, pp. 271-285, 2000.

[9] A. Arleo and W, Gerstner. "Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity." Biological Cybernetics 83, pp. 287-299, 2000.

[10] P. Gaussier, A. Revel, J. P. Banquet and V. Babeau. "From view cells and place cells to cognitive map learning: processing stages of the hippocampal system." Biological Cybernetics 86, pp. 15-28, 2002.

[11] A. Guazzelli, F. J. Corbacho, M. Bota, and M. A. Arbib. "Affordances, motivation, and the world graph theory." Adaptive Behavior: Special issue on biologically inspired models of navigation, vol. 6 (3/4), pp. 435-471, 1998.

[12] M. Milford and G. Wyeth. "Hippocampal Models for Simultaneous Localization and Mapping on an Autonomous Robot." Proceedings of the 2003 Australasian Conference on Robotics and Automation. Brisbane, Australia, 2003.

[13] D. S. Olton. "Memory Functions and the Hippocampus." In W. Seifert (Ed.), Neurobiology of the Hippocampus. New York, Academic Press, pp. 335 – 373, 1983.

[14] R. G. M. Morris. "An Attempt to Dissociate Spatial-Mapping and Working-Memory Theories of Hippocampal Funtion." In W. Seifert (Ed.), Neurobiology of the Hippocampus. New York, Academic Press, pp. 405 – 432, 1983.

[15] J. O'keefe. "Spatial memory within and without the hippocampal system." In W. Seifert (Ed.), Neurobiology of the Hippocampus. New York, Academic Press, pp. 375 – 403, 1983.

[16] D. O. Hebb. "The Organization of Behavior: A Neuropsychological Theory." Wiley-Interscience, New York, 1949.

[17] A. G. Barto. "Reinforcement learning." In M. A. Arbib (Eds.), Handbook of Brain Theory and Neural Networks, Cambridge, MA: MIT Press, pp. 804 – 809, 1995.

[18] J. J, Gibson. "The senses considered as perceptual systems." Allen and Unwin, 1966.

[19] W. Schultz, R. Romo, T. Ljungberg, J. Mirenowicz, J. R. Hollerman and A. Dickinson. "Reward-related signals carried by dopamine neurons." In J. C. Houk, J. L. Davis and D. G. Beiser (Eds.), Models of information processing in the basal ganglia. Cambridge, MA, The MIT Press, pp. 233-248, 1995.

[20] J. C. Houk, J. L. Adams, and A. G. Barto. "A model of how the basal ganglia generate and use neural signals that predict reinforcement." In J. C. Houk, J. L. Davis and D. G. Beiser (Eds.), Models of information processing in the basal ganglia. Cambridge, MA, The MIT Press, 1995.

[21] A. G. Barto. "Adaptive critics and the basal ganglia." In J. C. Houk, J. L. Davis and D. G. Beiser (Eds.), Models of information processing in the basal ganglia. Cambridge, MA, The MIT Press, 215-232, 1995.

[22] A. Barrera and A. Weitzenfeld. "Return of the Rat: Biologically-Inspired Robotic Exploration and Navigation." Proceedings of the 1st IEEE / RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob 2006). Pisa, Tuscany, Italy. February 20–22, 2006.

[23] A. Weitzenfeld, M. Arbib and A. Alexander. "The Neural Simulation Language." MIT Press, 2002.

[24] A. Barrera and A. Weitzenfeld. "Bioneural control model for robotic learning and mapping: robot experimentation results." ftp://ftp.itam.mx/pub/alfredo/ABarrera/Videos2005/.