



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN
INGENIERÍA

FACULTAD DE INGENIERÍA

**MODELO DE COGNICIÓN ESPACIAL Y NAVEGACIÓN
EN RATAS PARA CONTROLAR A UN ROBOT MÓVIL
AUTÓNOMO**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN INGENIERÍA

ELÉCTRICA – PROCESAMIENTO DIGITAL DE SEÑALES

P R E S E N T A :

MARÍA ALEJANDRA BARRERA RAMÍREZ

TUTOR: JESÚS SAVAGE CARMONA



2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

- Presidente: Dr. Tang Xu Yu
Facultad de Ingeniería
- Secretario: Dr. Weitzenfeld Ridel Alfredo
Asesor externo, ITAM
- Vocal: Dr. Savage Carmona Jesús
Facultad de Ingeniería
- 1^{er}. Suplente: Dr. Bermúdez Rattoni Federico
Instituto de Fisiología Celular
- 2^{do}. Suplente: Dr. Ramírez Amaya Víctor
Instituto de Neurobiología

Lugares donde se realizó la tesis:

FACULTAD DE INGENIERÍA
INSTITUTO DE NEUROBIOLOGÍA
INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

TUTOR DE TESIS:

DR. SAVAGE CARMONA JESÚS

FIRMA

CONTENIDO

I. INTRODUCCIÓN	1
I.1. Propósito de la Investigación.....	2
I.2. Organización de la Disertación	3
II. ANTECEDENTES DE INVESTIGACIÓN	5
II.1. Aprendizaje y Memoria Espacial en Ratas	5
II.1.1. Integración de Trayectoria.....	5
II.1.2. Señales Visuales en la Navegación	6
II.1.3. Mapas Cognitivos.....	7
II.1.4. La Motivación y el Sistema de Recompensas.....	8
II.2. Localización y Mapeo Simultáneos (SLAM)	9
II.2.1. Mapeo Topológico.....	9
II.2.2. Mapeo Métrico.....	10
II.2.3. Mapeo Híbrido.....	11
II.2.4. Asociación de Datos y Ambigüedad Perceptual.....	13
II.2.5. Evolución de los Algoritmos de SLAM	13
II.3. Trabajo Relacionado	14
II.3.1. El modelo de Burgess y colaboradores (1994)	14
II.3.2. El modelo de Redish y Touretzky (1997)	15
II.3.3. El modelo de Guazelli y colaboradores (1998).....	16
II.3.4. El modelo de Gaussier y colaboradores (2002)	17
II.3.5. El modelo de Arleo y colaboradores (2004)	18
II.3.6. El modelo de Milford y colaboradores (2006).....	19
III. MODELO DE COGNICIÓN ESPACIAL Y NAVEGACIÓN EN RATAS	20
III.1. Contexto Biológico.....	20
III.2. Motivación.....	23
III.3. Procesamiento de Affordances	23
III.4. Procesamiento de Información Kinestésica.....	25
III.5. Procesamiento de Información Visual.....	28
III.6. Representación y Reconocimiento de Lugares	30
III.7. Aprendizaje	34

III.8. Selección de Acción.....	35
IV. IMPLEMENTACIÓN DEL MODELO	41
IV.1. El Lenguaje de Simulación Neuronal	41
IV.2. Interfaz Modelo – Robot Virtual.....	43
IV.3. Interfaz Modelo – Robot Real	45
V. EXPERIMENTACIÓN.....	46
V.1. Discriminación: Aprendizaje – Desaprendizaje de Localidades de Recompensa.....	46
V.1.1. La Tarea Espacial de Discriminación Inversa de O’Keefe	47
V.1.2. Un Modelo de Navegación para Ratas Lesionadas.....	49
V.1.2.1. Procesamiento de Affordances.....	51
V.1.2.2. Aprendizaje.....	52
V.1.2.3. Selección de Acción	55
V.1.3. Resultados Experimentales	56
V.1.3.1. El Laberinto Tipo T y el Radial de 8 brazos con Ratas Sanas.....	56
V.1.3.2. El Laberinto Tipo T y el Radial de 8 brazos con Ratas Lesionadas.....	61
V.1.3.3. El Laberinto de Múltiples Ts con Ratas Sanas	64
V.2. Navegación Orientada a Metas	65
V.2.1. El Laberinto de Agua de Morris	66
V.2.2. Resultados Experimentales	67
V.3. Adaptación del Mapa Cognitivo ante Cambios Físicos	69
V.3.1. Intercambio de Landmarks	70
V.3.2. Eliminación de Landmarks.....	71
V.3.2.1. Removiendo L1.....	71
V.3.2.2. Removiendo L2.....	72
V.3.2.3. Removiendo L3.....	73
V.4. Reconocimiento de Lugares en Múltiples Direcciones	75
V.4.1. El Laberinto Cíclico: Resultados Experimentales con Ratas	76
V.4.1.1. Fase de Entrenamiento	76
V.4.1.2. Fase de Prueba	77
V.4.2. El Laberinto Cíclico: Resultados Experimentales con Robots	82
V.4.2.1. Fase de Habitación.....	82
V.4.2.2. Fase de Entrenamiento	83
V.4.2.3. Fase de Prueba	84
V.4.3. Análisis Conductual Comparativo	86
VI. DISCUSIÓN Y CONCLUSIONES.....	88
VI.1. Sobre los Resultados Experimentales	90

VI.2. Sobre la Contribución de la Investigación.....	92
VI.3. Sobre las Líneas Futuras.....	94
VII. APÉNDICE A: VALOR DE PARÁMETROS.....	97
VIII. GLOSARIO	99
IX. REFERENCIAS.....	101

RESUMEN

Actualmente no hay duda de que las ratas son capaces de generar una representación espacial del ambiente que exploran. Existe una gran cantidad de trabajo teórico y experimental al respecto, el cual considera dos teorías neurocientíficas fundamentales: la del mapa cognitivo de Tolman (1948) y la del hipocampo como un mapa cognitivo de O'Keefe y Nadel (1978). Específicamente, se han abordado aspectos relacionados con las estrategias de aprendizaje que las ratas llevan a cabo, la información de entrada al mapa cognitivo, la exhibición de conductas orientadas a metas, y las regiones cerebrales involucradas en todos estos procesos. Como resultado, han surgido valiosas hipótesis sobre los mecanismos que subyacen el aprendizaje y la memoria.

El objetivo de la investigación que aquí se presenta consistió en identificar algunas de estas hipótesis y utilizarlas para modelar computacionalmente las siguientes habilidades de las ratas: la exploración del ambiente, la construcción del mapa cognitivo, el aprendizaje y desaprendizaje de localidades meta, la adaptación del mapa ante cambios ambientales físicos, y la explotación del mismo durante la navegación hacia la localidad meta desde cualquier punto de partida. Estos procesos que las ratas ejecutan de manera natural, constituyen grandes retos para la programación de robots, por lo que investigadores en el campo de la robótica han tomado inspiración en la naturaleza de estos animales para atribuir a robots capacidades similares. Así lo hace este estudio, generando como resultado un modelo computacional que considera y extiende algunos de los trabajos existentes, y cuyo desempeño es demostrado en un robot móvil autónomo.

El modelo propuesto de cognición espacial y navegación en ratas comprende diversos módulos funcionales que capturan algunas propiedades de las estructuras del cerebro de la rata involucradas en el aprendizaje y la memoria. El modelo depende de la integración de información kinestésica y visual, y del uso de aprendizaje de Hebb (1949), para construir una representación espacial topológica-métrica durante la exploración del ambiente; considera el estado motivacional interno del animal al implementar aprendizaje por refuerzo a través de una arquitectura Actor-Crítico (Barto, 1995) para posibilitar el aprendizaje y desaprendizaje de localidades meta en su vecindad dentro del ambiente; y complementa dicha arquitectura de aprendizaje con un método de propagación del refuerzo hacia atrás a través de la trayectoria seguida por el animal durante cada ensayo de un experimento dado. En este método, la adaptación de las expectativas de recompensa asociadas con las direcciones apropiadas en las diversas localidades de la trayectoria, hace posible el aprendizaje y desaprendizaje de rutas en ambientes que incluyen varios puntos de decisión y donde la meta no es perceptible a distancia. Una vez explorado y aprendido el ambiente por parte de la rata, la explotación de las máximas expectativas de recompensa almacenadas en los nodos del mapa cognitivo le permiten alcanzar la meta siguiendo una ruta directa desde cualquier localidad inicial. Adicionalmente, el modelo propuesto implementa un proceso de adaptación en línea del mapa cognitivo ante situaciones en las que el ambiente cambia su configuración física durante la exploración por parte del animal.

Como resultado, el modelo determina las acciones motrices a ejecutar por la rata en cualquier localidad del ambiente, las cuales consisten de la próxima dirección de la cabeza de la rata, la rotación requerida para orientarse hacia tal dirección, y el desplazamiento a efectuar. Esta decisión del animal considera las posibles rotaciones a ejecutar a partir de la localidad y dirección actuales, la curiosidad por efectuar rotaciones aún no exploradas, y las expectativas de máxima recompensa.

El modelo de cognición espacial y navegación en ratas fue diseñado e implementado utilizando el Lenguaje de Simulación Neuronal NSL (Weitzenfeld et al., 2002), y es capaz de interactuar con un ambiente virtual o real a través de dos módulos externos de procesamiento visual y de control motriz. A nivel simulado, el modelo utiliza las imágenes provenientes del mundo virtual, y controla la rotación y el desplazamiento del robot virtual dentro de su mundo. Por otro lado, al experimentar con un robot real, el modelo considera las

imágenes tomadas por el robot con su cámara local, y lo controla rotándolo y desplazándolo dentro del ambiente físicamente construido.

Cuatro fueron los aspectos probados durante la experimentación con el robot: (i) el aprendizaje y desaprendizaje de localidades de recompensa, (ii) la navegación orientada a metas, (iii) la adaptación del mapa cognitivo ante cambios físicos en el ambiente, y (iv) el reconocimiento de lugares en múltiples direcciones. Para el caso (i), la disertación describe los resultados derivados de implementar con robots una de las tareas espaciales clásicas de discriminación inversa con ratas sanas y ratas con hipocampo lesionado en un laberinto tipo T y en uno radial de 8 brazos (O'Keefe, 1983), y muestra la similitud conductual entre el robot y las ratas.

Considerando el aspecto (ii), se presentan los resultados obtenidos al experimentar con el robot en una tarea de protocolo similar a uno de los experimentos espaciales clásicos de navegación orientada a metas (Morris, 1981) empleando un laberinto formado por múltiples Ts rodeado por claves visuales o landmarks, donde fue posible confirmar en el caso de los robots "sanos", como lo hizo Morris con las ratas sanas, que para encontrar una meta "escondida" en el laberinto independientemente de la localidad de partida, los robots requiere explotar su mapa cognitivo reconociendo y utilizando las expectativas de recompensa que les hacen navegar hacia la localidad objetivo exitosamente.

En relación al caso (iii), la disertación discute los resultados de probar al robot en el laberinto de múltiples Ts rodeado por landmarks alterando la disposición física de estas claves visuales después de haberlo entrenado para encontrar la meta. En particular, se experimentó separadamente con dos escenarios: el intercambio y la eliminación de landmarks, y fue posible verificar que el robot, en tiempo real, es capaz de adaptar el mapa cognitivo agregando y/o eliminando nodos para reflejar los cambios percibidos.

Por último, con respecto al caso (iv), la disertación describe los resultados experimentales obtenidos con ratas y con robots resolviendo la misma tarea espacial dentro de un laberinto cíclico, y presenta el correspondiente análisis conductual comparativo. La exploración exhaustiva efectuada durante la fase de habituación en este laberinto permitió al robot representar en el mapa cognitivo todas las localidades visitadas en todas las orientaciones posibles, agregar la información relacionada con expectativas de recompensa durante el entrenamiento, reconocer los lugares en todos los ensayos de prueba sin agregar nuevos nodos al mapa, y explotar la información almacenada en los nodos existentes para alcanzar la localidad meta ejecutando las rotaciones asociadas con las máximas expectativas de recompensa. En cuanto a los resultados derivados del experimento con ratas en el laberinto cíclico, éstos no revelaron un desempeño eficiente en todas las situaciones probadas. De hecho, la rata reportó casos de trayectorias indirectas que implicaron pasar dos veces por los mismos lugares aún después de haber aprendido la tarea.

Desde la perspectiva robótica, la contribución del presente trabajo consistió en abordar la inequidad entre mapeo y explotación del mapa que existe en la literatura sobre localización y mapeo simultáneos (SLAM), así como la falta de investigación sobre desaprendizaje, tomando inspiración de tales habilidades en las ratas. El interés particular del trabajo radicó en entender los mecanismos que subyacen al proceso de cognición espacial en las ratas, con el fin de incorporar datos fisiológicos relativos a dicho proceso en una arquitectura robótica que fuera evaluada desde una perspectiva conductual a través de la comparación con resultados biológicos. De esta forma, se considera que la investigación contribuye a la neurociencia experimental ofreciendo el sistema propuesto como una herramienta para probar con robots nuevas hipótesis que pudieran extender el conocimiento actual sobre aprendizaje y memoria en ratas.

Finalmente, dentro del campo de la robótica, muchos aspectos deben ser abordados antes de situar el sistema propuesto dentro de una escala de desempeño o compararlo con los resultados obtenidos por enfoques particulares de SLAM; y desde la perspectiva biológica, el modelo básico de cognición espacial y navegación en ratas debe ser sofisticado incorporando conocimiento adicional sobre aprendizaje y memoria, con la finalidad de habilitarlo para proveer retroalimentación valiosa a la neurociencia experimental. Así, la disertación concluye discutiendo los aspectos a tratar como trabajo futuro tanto en el ámbito de la robótica como de la biología.

ABSTRACT

It is well known currently that rats are able to generate a spatial representation of any explored environment. Extensive related theoretical and experimental work considers two fundamental neuroscientific theories: the cognitive map by Tolman (1948), and the hippocampus as a cognitive map by O'Keefe and Nadel (1978). Specifically, the following aspects have been addressed among others: the learning strategies employed by rats, the input to the cognitive map, the exhibition of goal-oriented behaviors, and the brain areas involved in these processes. As a result, valuable hypotheses have come up to explain the mechanisms underlying learning and memory.

The purpose of this research was to identify some of those hypotheses and employ them to computationally model the following rats' abilities: the exploration of the environment, the construction of the cognitive map, the learning and unlearning of goal locations, the adaptation of the map to physical changes in the environment, and the map exploitation during navigation to the goal location from any given point of departure. All these processes that rats perform naturally constitute big challenges for robots. Therefore, researchers on the field of robotics have taken inspiration from rats' nature to endow robots with similar capabilities. This is what our study does, producing as a result a computational model that considers and extends existent work, and which performance is demonstrated on an autonomous mobile robot.

The proposed model of spatial cognition and navigation in rats includes several functional modules that capture some properties of the rat brain structures involved in learning and memory. The model (i) relies on the integration of kinesthetic and visual information and on the use of Hebbian learning (1949) to build a topological-metric spatial representation during the environment exploration; (ii) considers the internal motivational state of the animal to implement reinforcement learning through an Actor-Critic architecture (Barto, 1995) to enable the learning and unlearning of goal locations in their neighborhood within the environment; and (iii) complements this learning architecture with a reinforcement back-propagation method applied to the route followed by the animal during every trial in a given experiment. In this method, the adaptation of reward expectations associated to appropriate directions at the different locations of the route enables the learning and unlearning of routes within environments including several choice points, where the goal is perceptible only at its location. Once the rat has explored and learnt the environment, the exploitation of maximum expectations of reward stored in the nodes of the cognitive map allows the animal to reach the goal following a direct route from any given departure location. Additionally, the model implements an on-line adaptation process of the cognitive map to situations where changes in the environment physical configuration occur during exploration.

As output, the model determines the motor actions to be performed by the rat at any environmental location, which include the next rat's head direction, the rotation required to point to that direction, and the next motion displacement. The motor decision of the animal is taken by considering all possible rotations to execute from current location and orientation, the curiosity to execute rotations not yet explored, and the expectations of maximum reward.

The model of spatial cognition and navigation in rats was designed and implemented by using the Neural Simulation Language (Weitzenfeld et al., 2002), and it is able to interact with a virtual or real environment through two external visual processing and motor control modules. During simulations, the model employs images taken from the virtual world, and controls the rotation and translation of the virtual robot within its world. While experimenting with a real robot, on the other hand, the model considers images taken by the robot with its local camera, and controls the robot by rotating and translating it within the physically built environment.

Four aspects were tested while experimenting with the robot: (i) the learning and unlearning of reward locations, (ii) the goal-oriented navigation, (iii) the adaptation of the cognitive map to physical changes in the environment, and (iv) the place recognition in multiple directions. Considering case (i), the dissertation

describes results derived from implementing with robots the classical spatial reversal task carried out with normal and hippocampal-lesioned rats in a T-maze and in an 8-arm radial maze (O'Keefe, 1983), and shows the behavioral similitude between the robot and the rats.

In relation to aspect (ii), the dissertation presents results from experimenting with the robot in a spatial task that involves a similar protocol used in a classical goal-oriented navigation experiment (Morris, 1981), employing in our case a multiple T-maze surrounded by landmarks. It could be confirmed in the situation of "normal" robots, as Morris did with normal rats, that in order to find a "hidden" target in the maze from any given departure location, robots need to exploit their cognitive map by recognizing and using the expectations of reward that make them navigate towards the goal location successfully.

In case (iii), the dissertation discusses results from testing the robot in the multiple T-maze surrounded by landmarks when changing the physical configuration of landmarks after having trained the robot to find the goal. Particularly, we experimented separately with two scenarios: the interchange and elimination of landmarks, and we could verify that the robot is able to adapt the cognitive map in real time by adding and/or removing nodes to represent the perceived changes.

In reference to aspect (iv), the dissertation describes experimental results obtained with rats and robots independently solving the same spatial task within a cyclic maze, and presents the corresponding comparative behavioral analysis. The exhaustive exploration carried out during the habituation phase in this maze allowed the robot to represent all visited locations in all possible orientations within the cognitive map, add information related to reward expectations during training, recognize the environmental locations during all testing trials without adding new nodes to the map, and exploit the information stored in existent nodes to reach the goal by executing rotations corresponding to maximum expectations of reward. Considering the results derived from the experiment with rats, they did not reveal an efficient performance in all situations tested. In fact, the rat reported cases of indirect routes, where some places were visited twice after having learnt the task.

From a robotics perspective, the contribution of this work consisted on addressing the imbalance between mapping and map exploitation currently existent in the SLAM literature, as well as the lack of unlearning research, by taking inspiration from these abilities in rats. The specific concern of the work was to understand the underlying mechanisms of rats' spatial cognition in order to incorporate relative physiological data in a robotic architecture to be evaluated from a behavioral perspective through the comparison with biological results. In this way, the study contributes to experimental neuroscience by providing the proposed system as a tool to test with robots new hypotheses that might extend the current knowledge on learning and memory in rodents.

Finally, within the field of robotics, many aspects should be addressed before placing the proposed system within a performance scale or comparing it with specific SLAM approaches; and from a biological perspective, the basic model of spatial cognition and navigation in rats must be sophisticated by incorporating additional knowledge on learning and memory in order to enable it to provide valuable feedback to experimental neuroscience. In this way, the dissertation concludes discussing the aspects to address as future work in both research areas: robotics and biology.

I. INTRODUCCIÓN

La tecnología siempre ha jugado un papel crucial en los intentos por entender la mente y el cuerpo humanos. En 1748, La Mettrie publicó *L'Homme Machine* y sugirió la posibilidad de que algún día se construyera un hombre mecánico que pudiera hablar, aún cuando los autómatas de aquellos tiempos eran incapaces de adaptarse a circunstancias cambiantes (Arbib, 1987). En 1878, Bernard observó que los procesos fisiológicos ocurrientes en el cuerpo humano formaban, frecuentemente, ciclos causa-efecto, los cuales podían contrarrestar los disturbios que sucedieran en variables como la temperatura, la presión sanguínea y el nivel de glucosa en la sangre. Tal observación involucraba importantes conceptos como la retroalimentación negativa y la estabilidad de los sistemas.

El año 1943 resultó clave para la conjunción de las nociones de mecanismos de control y autómatas inteligentes (Arbib, 1987). Craik, en su ensayo *The Nature of Explanation*, apreció al sistema nervioso como una máquina calculadora capaz de modelar los eventos del mundo externo, sugiriendo que tal capacidad constituía la característica básica del pensamiento. Por otro lado, Rosenblueth, Wiener y Bigelow ratificaron la observación de Bernard al encontrar evidente la existencia del proceso de retroalimentación dentro del sistema nervioso; esto es, se percataron de que la información externa llegaba al sistema nervioso a través de receptores, del sistema nervioso se transmitía a los músculos y a su entorno, y partiendo de aquí, regresaba al sistema nervioso nuevamente mediante los receptores. También en ese mismo año, McCulloch y Pitts estudiaron “la fisiología de lo computable”, propusieron un modelo formal de neurona concibiéndola como una unidad lógica de umbral, y demostraron que la funcionalidad de una red de estas unidades podía compararse, desde la perspectiva computacional, con la ofrecida por los autómatas definidos por Alan Turing a mediados de los años 30.

Con base en investigaciones como las publicadas en 1943, Wiener, en 1948, definió el término Cibernética como el estudio del control y la comunicación en animales y máquinas. Tal enfoque floreció, pero más tarde se disolvió, ya que para los años 60, prácticamente se había perdido el interés que fundamentaba los trabajos científicos relacionados con la comparación entre las características estructurales y funcionales del hombre y la máquina. Únicamente importaba encontrar la manera en que las computadoras debían ser programadas para exhibir aspectos característicos de la inteligencia humana con el afán de mejorar el desempeño de las máquinas, mas no de conocer los mecanismos por los que los humanos son inteligentes. En otras palabras, se perseguía el estudio de la Inteligencia Artificial, considerándola como un fin en sí misma (Arbib, 1989).

Muchos eran los dominados por la metáfora “los humanos son máquinas”; sin embargo, ésta llegó a ser interpretada desde dos perspectivas distintas. El enfoque defendido por la Inteligencia Artificial alude a la programación de las computadoras para mostrar “comportamientos inteligentes”, sin intentar correlacionar las estructuras del programa con estructuras cerebrales. Por otro lado, según la perspectiva de la Teoría del Cerebro, se utilizan la simulación por computadora y el análisis matemático para exhibir cómo las interacciones entre los componentes del cerebro pueden sustentar funciones como la percepción, la memoria y el control del movimiento.

Dentro del campo de la Teoría del Cerebro se han desarrollado dos áreas de estudio diferentes aunque interactivas: la Neurociencia Computacional, que tiene como meta desarrollar modelos computacionales sobre la estructura y funcionamiento del cerebro para explicar datos biológicos; y la Biorobótica, que aprovecha el

entendimiento del cerebro a través de los modelos computacionales para construir sistemas que posibiliten a los robots manifestar conductas “inteligentes”.

La inspiración en el cerebro acerca al diseño de robots con capacidades de aprendizaje y de adaptación en un entorno cambiante; de aquí que la interacción entre las Neurociencias y la Robótica merezca ser estudiada. El marco de referencia para dicho estudio se presenta esquematizado en la Figura I.1.

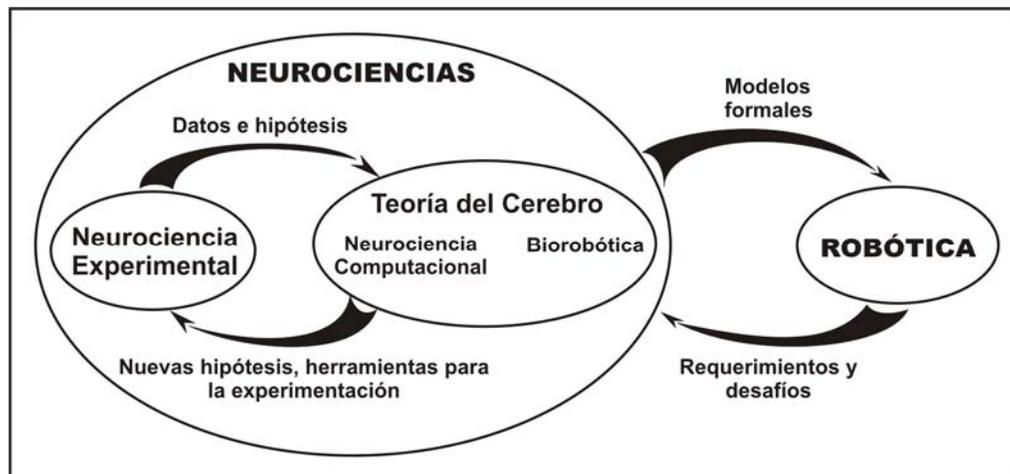


Figura I.1. Marco de referencia para la interacción entre las Neurociencias y la Robótica.

El desarrollo de modelos computacionales implica un trabajo interdisciplinario en el que prevalece el intercambio de conceptos entre teóricos y experimentalistas. El término Neurociencia Computacional proviene, precisamente, de la necesidad de emplear conceptos y metodologías computacionales para estudiar los datos neurocientíficos, considerando las hipótesis que sustentan la experimentación. La Robótica, por otro lado, puede considerar los modelos computacionales derivados de las Neurociencias en el diseño de robots, y retroalimentar su labor indicando los desafíos que interesa satisfacer para la construcción de robots “inteligentes”.

Cuando los modelos logran generar datos que pueden ser validados por los resultados experimentales, es factible que sean empleados por los neurocientíficos experimentalistas como herramientas para simular experimentos con robots que, de producir resultados interesantes, podrán guiar los siguientes experimentos que efectúen en la realidad. Posteriormente, los resultados empíricos, ya sea que coincidan o difieran de los derivados artificialmente, servirán a los neurocientíficos para plantear nuevas hipótesis o para ajustar las que se encuentran incrustadas en los modelos desarrollados. De esta forma se cierra el ciclo Teoría – Experimento que permite introducir el trabajo aquí propuesto.

I.1. PROPÓSITO DE LA INVESTIGACIÓN

La supervivencia de muchas especies depende de su habilidad para representar y utilizar conocimiento espacial que resulta fundamental no sólo para escapar de situaciones peligrosas, sino también para regresar a lugares que proveen comida y agua.

Actualmente no hay duda de que las ratas son capaces de generar una representación espacial del ambiente que exploran. Existe una gran cantidad de trabajo teórico y experimental al respecto, el cual considera dos teorías neurocientíficas fundamentales: la del mapa cognitivo de Tolman (1948) y la del hipocampo como un mapa cognitivo de O’Keefe y Nadel (1978). Específicamente, se han abordado aspectos

relacionados con las estrategias de aprendizaje que las ratas llevan a cabo, la información de entrada al mapa cognitivo, la exhibición de conductas orientadas a metas, y las regiones cerebrales involucradas en todos estos procesos, entre otros. Como resultado, han surgido valiosas hipótesis sobre los mecanismos que subyacen el aprendizaje y la memoria, aunque no todas las respuestas están dadas.

El objetivo de la investigación que aquí se presenta consiste en identificar algunas de estas hipótesis y utilizarlas para modelar computacionalmente las siguientes habilidades de las ratas: la exploración del ambiente, la construcción del mapa cognitivo, el aprendizaje y desaprendizaje de localidades meta, la adaptación del mapa ante cambios ambientales físicos, y la explotación del mismo durante la navegación hacia la localidad meta desde cualquier punto de partida. Estos procesos que las ratas ejecutan de manera natural, constituyen grandes retos para la programación de robots, por lo que investigadores en el campo de la robótica han tomado inspiración en la naturaleza de estos animales para atribuir a robots capacidades similares. Así lo hace el presente estudio, generando como resultado un modelo computacional que considera y extiende algunos de los trabajos existentes, y cuyo desempeño es demostrado en un robot móvil autónomo.

De acuerdo con el marco de referencia ilustrado en la Figura I.1, este estudio puede ubicarse en el ámbito de la Teoría del Cerebro, y sus metas impactan sobre la Neurociencia Experimental y la Robótica.

En particular, el modelo formal incorpora hipótesis y datos empíricos relativos a la cognición espacial y navegación en ratas, esperando sugerir nuevas hipótesis neurocientíficas, promover el diseño de nuevos experimentos que permitan incrementar el entendimiento de la conducta de las ratas a partir de los resultados derivados de la experimentación con robots, y proveer a los neurocientíficos una herramienta para probar nuevas hipótesis con robots antes de hacerlo con ratas. Sin duda, el modelo formal habrá de ser enriquecido continuamente para concretar y consolidar sus aportaciones a la Neurociencia Experimental.

Por otro lado, el modelo resultante de este estudio se sitúa en el espacio que existe en la investigación sobre localización y mapeo robóticos, en donde la mayoría de los algoritmos propuestos hasta ahora abordan las representaciones espaciales generadas por los robots, mas no la explotación de éstas para navegar directamente a localidades meta designadas, y tampoco la adaptación de las mismas como resultado del desaprendizaje de las localidades meta.

I.2. ORGANIZACIÓN DE LA DISERTACIÓN

Al terminar este primer capítulo de la disertación, el resto queda organizado en diversas secciones temáticas. Así, el segundo capítulo presenta los antecedentes de investigación referentes a los rubros que enmarcan el trabajo. En particular, se describen los mecanismos neurofisiológicos y conductuales que subyacen el aprendizaje y la memoria espacial en ratas, se plantea el estado del arte en los algoritmos de localización y mapeo simultáneos (SLAM) aplicados a robots móviles, y se introducen algunos de los esfuerzos más importantes documentados en los últimos años acerca de modelos de cognición espacial y navegación en ratas aplicados a robots móviles.

El tercer capítulo presenta el modelo propuesto de cognición espacial y navegación en ratas, explicando el contexto biológico que lo subyace, y describiendo detalladamente cada uno de los módulos funcionales que lo componen.

El cuarto capítulo discute los aspectos de implementación del modelo propuesto, la cual involucra tanto el diseño de sus módulos, como la definición de las interfaces que le permiten interactuar con el ambiente. En concreto, el capítulo incluye tres secciones que describen el uso del Lenguaje de Simulación Neuronal NSL en el diseño de los componentes del modelo, la interacción del modelo con un robot virtual, y la interfaz entre

el modelo y el robot real. Estos dos últimos apartados se enfocan a la especificación del procesamiento visual y del control motriz del robot.

El quinto capítulo aborda, a través de cuatro secciones, los aspectos probados durante la experimentación con robots. Así, una primera sección se enfoca en los procesos de aprendizaje y desaprendizaje de localidades de recompensa, introduciendo conceptos de cognición espacial relacionados con la discriminación, la discriminación inversa, la extinción de recompensas y la persistencia, como antecedente para presentar una de las tareas espaciales clásicas de discriminación inversa con ratas sanas y ratas con hipocampo lesionado. Posteriormente, la sección exhibe la propuesta de una versión adaptada del modelo de navegación para el caso de ratas lesionadas, y describe los resultados derivados de implementar dicha tarea de discriminación inversa con robots controlados por el modelo de cognición espacial originalmente propuesto, y con robots dirigidos por el modelo de navegación adaptado.

Por otro lado, la segunda sección aborda la navegación orientada a metas, presentando los resultados obtenidos al experimentar con el robot en una tarea espacial en donde es factible alcanzar la localidad meta desde cualquier otra localidad del laberinto aprendido. La tercera sección describe el proceso de adaptación del mapa cognitivo para reflejar cambios en la configuración física de un laberinto, así como los resultados obtenidos de la experimentación con el robot al respecto. Finalmente, la cuarta sección se enfoca al proceso de reconocimiento de lugares que llevan a cabo tanto las ratas como el robot durante la navegación orientada a metas. En particular, se describen los resultados experimentales obtenidos con ratas y con robots resolviendo la misma tarea espacial dentro de un laberinto cíclico, y se presenta el correspondiente análisis conductual comparativo.

Por último, el sexto capítulo de la disertación presenta la discusión final concerniente a los resultados experimentales obtenidos del proyecto de investigación, a su contribución en los campos de la Robótica y de las Neurociencias, y a las líneas de investigación que serán abordadas como metas próximas.

II. ANTECEDENTES DE INVESTIGACIÓN

Este capítulo presenta un panorama general de los aspectos teóricos que enmarcan el trabajo y que han sido estudiados con el fin de soportar el desarrollo del modelo de cognición espacial y navegación en ratas, así como la generación de los resultados con un robot móvil.

El capítulo queda dividido en tres apartados: el primero describe los mecanismos neurofisiológicos y conductuales que subyacen el aprendizaje y la memoria espacial en ratas, el segundo plantea el estado del arte en los algoritmos de localización y mapeo simultáneos (SLAM) aplicados a robots móviles, y finalmente, el tercer apartado introduce algunos de los esfuerzos más importantes documentados en los últimos años acerca de modelos de cognición espacial y navegación en ratas aplicados a robots móviles.

II.1. APRENDIZAJE Y MEMORIA ESPACIAL EN RATAS

La posición en el espacio es parte de la información fundamental que los animales y las personas aprenden. La ubicación de la comida, el agua y el hogar debe ser recordada para poder regresar, y la localización de objetos dañinos y predadores debe ser recordada para que ambos puedan ser evitados. Además de la memoria de la posición de objetos o lugares, la memoria para moverse a través del espacio resulta igualmente importante para llegar a un punto B a partir de un punto A en el que se encuentre localizado el individuo. Se utiliza el término **cognición espacial** para referir el proceso de codificación y almacenamiento de información espacial, la cual permite que la orientación hacia los objetos y el movimiento a través del espacio sean exitosos (Roberts, 1998).

Los procesos espaciales de aprendizaje y memoria en las ratas involucran la interacción de diversas estructuras cerebrales para manifestar habilidades relacionadas con la integración de la trayectoria recorrida, el uso de señales visuales en la orientación, la generación de representaciones espaciales, el manejo de recompensas, y la navegación dentro del ambiente. Las secciones que a continuación se presentan describen tales habilidades.

II.1.1. Integración de Trayectoria

El proceso a través del cual un sujeto en movimiento es capaz de mantener un registro de su posición en relación al lugar de partida es llamado **integración de trayectoria** (Roberts, 1998). Este nombre refleja la adición sucesiva de pequeños incrementos de movimiento dentro de una representación continuamente actualiza de la dirección y la distancia recorrida desde el punto de partida (Mittelstaedt y Mittelstaedt, 1982).

En ausencia de visión, la integración de trayectoria involucra dos sistemas receptores. Por un lado, los órganos vestibulares en los canales semicirculares del oído interno responden ante la aceleración y la velocidad de movimiento a la largo de un camino recto, y ante la velocidad de rotación cuando se efectúa un

giro. Por otro lado, información adicional es provista por la retroalimentación kinestésica procedente de los músculos cuando se realizan movimientos en el espacio. Además, la habilidad del sujeto para estimar el tiempo también es importante, pues conociendo la velocidad de movimiento y el tiempo que tomó el desplazamiento entre dos puntos, es factible estimar la distancia recorrida. Entonces, el proceso de integración de trayectoria es posible al mantener un registro de la distancia viajada a lo largo de un camino recto y de los ángulos en que se ejecutaron las rotaciones. Sin embargo, estos cálculos de rotaciones y traslaciones son estimaciones que involucran un nivel de error, el cual se acumula a medida que se efectúan más giros y se recorren distancias mayores.

Durante la exploración de un nuevo ambiente, la integración de trayectoria opera en diversas especies de vertebrados e invertebrados que manejan puntos de partida fijos. Tal es el caso de las hormigas, las arañas, los roedores, los perros y los humanos, entre otros. En particular, el comportamiento de integración de trayectoria en roedores ha sido estudiado por investigadores como Mittelstaedt y Mittelstaedt (1982), McNaughton y colaboradores (1996), Etienne (1998), y Etienne y Jeffrey (2004).

Mittelstaedt y Mittelstaedt (1982), y Etienne (1998) han mostrado que los roedores son capaces de ejecutar un camino directo de regreso a la localidad de partida en ambientes con claves controladas incluso después de haber recorrido una trayectoria compleja desde ahí. Cuando las claves relevantes son movidas, los animales las ignoran y regresan al punto de partida, lo cual sugiere que efectivamente llevan un registro de su posición relativa a la localidad de inicio a través del proceso de integración de trayectoria.

II.1.2. Señales Visuales en la Navegación

Dado que el proceso de integración de trayectoria se degrada por la acumulación de errores que ocurren al calcular automática y continuamente las traslaciones y rotaciones ejecutadas por el animal cada vez que se desplaza, el sistema de navegación apela a referencias externas típicamente visuales cuando éstas se encuentran disponibles para el animal.

Estudios conductuales y registros neuronales en ratas han mostrado que las señales visuales, también referidas como **landmarks**, controlan la navegación sólo si la rata las percibe como estables (Jeffery, 1998; Knierim et al., 1996), y de manera opuesta, el control visual se debilita después de experimentar inestabilidad (Knierim et al., 1996).

Las ratas también pueden utilizar estímulos no visuales como referencias externas y asociarlas con localidades o direcciones particulares. Una vez establecidas dichas asociaciones, las ratas muestran una gran flexibilidad en el uso de señales espaciales de diferentes modalidades sensoriales (Etienne, 2004). Sin embargo, en condiciones donde prevalecen diferentes tipos de referencias, las ratas muestran una jerarquía de preferencias: la visión sobre el olfato, y el olfato sobre la locomoción (Maaswinkel y Whishaw, 1999).

En condiciones naturales, las referencias externas y las señales derivadas del propio movimiento usualmente cooperan más que competir entre sí. De esta forma, el proceso de integración de trayectoria puede ser reiniciado mediante landmarks estables para guiar la navegación de manera más precisa cuando los landmarks dejan de estar disponibles. La reinicialización implica que el animal establezca su localización y orientación actuales a través de la percepción de las referencias externas, y que actualice el estado de su integrador de trayectoria de acuerdo con la posición del punto de partida. Este proceso ha sido demostrado en el comportamiento de retorno al punto de partida en hamsters cuando enfrentan situaciones de conflicto entre landmarks visuales y la integración de trayectoria (Etienne et al., 2000).

II.1.3. Mapas Cognitivos

Un estudio decisivo en la evolución del entendimiento acerca de los mecanismos cerebrales involucrados en el procesamiento de la información espacial fue sin duda el análisis que O'Keefe y Nadel (1978) llevaron a cabo sobre las propiedades neurofisiológicas y neuropsicológicas del hipocampo, descubriendo la participación crucial de este sistema en (i) el desarrollo de representaciones internas de alto nivel sobre las relaciones espaciales aloclínicas, esto es, basadas en una referencia absoluta y no en la posición actual del animal, y en (ii) el aprendizaje para resolver problemas que requieren la memoria de tales relaciones. La Figura II.1 muestra la anatomía del hipocampo de la rata en forma esquemática.

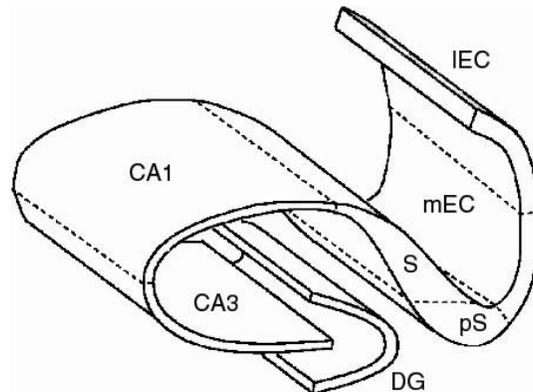


Figura II.1. Corte esquemático perpendicular al eje longitudinal del hipocampo de la rata. EC: corteza entorhinal (mEC: EC medial, IEC: EC lateral); S: subículo; pS: presubículo y parasubículo; DG: giro dentado; CA1-CA3: áreas CA1-CA3. Adaptado de (Burgess y O'Keefe, 2002).

El sustrato neuronal de dichas representaciones espaciales internas fue el descubrimiento previo que hicieron O'Keefe y Dostrovsky (1971) a partir de registros individuales de las células piramidales encontradas en las áreas CA3 y CA1 del hipocampo. A tales neuronas se les llamó **células de lugar** (*place cells*), pues muestran una alta correlación entre su tasa de disparo y la localización del animal en el ambiente. Específicamente, cada una de estas células exhiben patrones de activación tipo Gaussiano a una tasa elevada cuando el animal se encuentra en un área continua y compacta, la cual es definida como el **campo de lugar** (*place field*) de la célula (O'Keefe y Dostrovsky, 1971). La actividad derivada de la población de células de lugar codifica en conjunto la ubicación actual del animal dentro de un ambiente familiar, y es dicha actividad global la que se almacena en la representación espacial interna referida por Tolman (1948) como **mapa cognitivo**.

La actividad de las células de lugar parece ser dependiente de la disposición de señales visuales en el ambiente, ya que la rotación de tales claves provoca una rotación correspondiente en los campos de lugar (McNaughton et al., 1994(1)). Sin embargo, las células de lugar mantienen sus campos de lugar cuando algunas de las claves visuales se retiran del ambiente (O'Keefe y Conway, 1978), e incluso continúan respondiendo en la oscuridad (Quirk et al., 1990), así que actualmente se asume que la respuesta de las células de lugar se deriva de combinar información kinestésica con información de señales visuales (Jeffery y O'Keefe, 1999).

Según diversos investigadores (McNaughton et al., 1996; Etienne, 1998), el proceso de integración de trayectoria puede jugar el papel principal en la construcción del mapa cognitivo. En efecto, cuando las ratas exploran un nuevo ambiente, centran su excursión en el punto de partida y utilizan la integración de trayectoria continuamente para actualizar la relación entre su posición actual y el punto de referencia. En esta situación inicial, los campos de lugar se establecen rápidamente a través de la información kinestésica pues las señales ambientales no han sido aprendidas, pero conforme la exploración progresa, las ratas comienzan a asociar las diversas localidades relevantes del ambiente con su percepción actual de landmarks.

De acuerdo con Poucet (1993), el mapa cognitivo se genera mediante la adquisición de información topológica y métrica del espacio. La información topológica se refiere al conocimiento de las relaciones espaciales entre localidades u objetos, mientras que la información métrica se relaciona con información cuantitativa acerca de direcciones angulares específicas y distancias entre localidades u objetos. Poucet sugiere que los mapas cognitivos iniciales son mapas topológicos básicos que se van refinando con la adquisición de información métrica a medida que se explora el ambiente.

Un animal empieza por formar un mapa cognitivo de una región, construyendo gráficas locales que especifican las relaciones entre los lugares de un área. Poucet indica que la integración de vistas locales es importante para aprender acerca de un lugar. Si un animal permanece en un lugar y rota su cabeza y su cuerpo gradualmente hasta completar 360°, las vistas locales sucesivas son integradas para componer una representación completa de ese lugar. La vista local que el animal percibe al visitar nuevamente ese lugar debe reactivar el recuerdo entero del mismo. Por otro lado, cuando el animal viaja entre lugares, se establecen en el mapa relaciones acerca de qué lugar conduce a qué otro. A medida que se incrementa la exploración, se desarrolla información métrica; en particular, el animal aprende las distancias y las direcciones entre lugares. Finalmente, el animal forma un mapa global del área, el cual, según Poucet, es el resultado de integrar las gráficas locales, ligando una con otra a través de un lugar común y utilizando el conocimiento que el animal tiene acerca de la dirección y la distancia entre lugares. Dicho mapa global le permitirá al animal determinar la dirección para llegar de un punto a cualquier otro. Sin embargo, para lograrlo requiere establecer una dirección de referencia, como señales celestiales, una línea de árboles, las paredes de un laboratorio o una puerta. Poucet propone que los ejes principales de viaje en el ambiente pueden servir como referencia.

II.1.4. La Motivación y el Sistema de Recompensas

El comportamiento motivado usualmente está orientado a una meta, la cual puede asociarse en los animales con una necesidad como el hambre o la sed. Sin embargo, la motivación no ocurre solamente ante la sensación instintiva de hambre, sino que también puede ser estimulada por factores externos llamados **incentivos** como la percepción o el olor de la comida, por ejemplo. De acuerdo con Toates (1986), la motivación se deriva tanto del estado interno del individuo como de la presencia de incentivos.

El hipotálamo es considerado como el área principal del cerebro donde la información relacionada con el estado interno de la rata es combinada con incentivos (Risold et al., 1997). Específicamente, se asume que la búsqueda y la ingestión de comida son actividades controladas por el hipotálamo lateral (Kelley, 2004), el cual determina la recompensa primaria o inmediata que dichas actividades producen en el animal.

Además del hipotálamo, existen otras estructuras cerebrales dedicadas a la extracción de información relacionada con recompensas a partir de estímulos ambientales, y al uso de tal información en la generación de comportamientos orientados a metas. Se trata del estriado, uno de los núcleos que pertenecen a los ganglios basales, y de la corteza frontal (Schultz et al., 1998, 2000).

Las neuronas dopaminérgicas en los ganglios basales responden ante recompensas primarias derivadas de estímulos impredecibles, y ante estímulos condicionados asociados con recompensas. La función principal de estas neuronas es detectar errores en la predicción de recompensas. La señal de error producida por la población de neuronas es proyectada como una señal de refuerzo a un gran número de neuronas del estriado y de la corteza frontal modificando fuerzas sinápticas para posibilitar el aprendizaje de estímulos que conducen a una meta.

Las neuronas del estriado, por otro lado, se activan cuando el animal espera recompensas predecibles. Durante el aprendizaje de alguna tarea, tales respuestas ocurren inicialmente en ensayos recompensados y no recompensados, pero se limitan después únicamente a ensayos recompensados, lo cual ocurre en paralelo con la adaptación que hace el animal de sus expectativas de recompensa inferidas a partir de su

comportamiento (Schultz et al., 1998). El procesamiento de recompensas que lleva a cabo el estriado se considera más complejo que el observado en las neuronas dopaminérgicas, pues incorpora información conductual y provee información más detallada acerca de la predictibilidad (Hollerman et al., 2000).

La activación de las neuronas en la corteza frontal no depende tanto de la integración de información conductual y de recompensas como de la distinción entre diversas recompensas. En situaciones complejas que presentan más de una recompensa, las neuronas frontales proveen las señales que le permiten al animal seleccionar el comportamiento que lo conduzca a la meta más deseable. Este aspecto de elegir entre recompensas refleja un nivel de procesamiento no observado en las neuronas dopaminérgicas ni en las estriatales.

Así, diferentes estructuras procesan la información de recompensas de formas distintas. Mientras que las neuronas dopaminérgicas emiten una señal de enseñanza acerca de una recompensa sin importar de qué recompensa se trate, las neuronas del estriado adaptan las expectativas ante nuevas situaciones de recompensa, y las neuronas frontales procesan la naturaleza específica de las recompensas. Sin embargo, las tres señales cooperan para utilizar la información de recompensas durante el comportamiento orientado a metas (Hollerman et al., 2000).

II.2. LOCALIZACIÓN Y MAPEO SIMULTÁNEOS (SLAM)

La localización y mapeo simultáneos (SLAM) refiere el problema de un robot móvil construyendo un mapa de su ambiente mientras simultáneamente se localiza dentro de dicho mapa (Hähnel et al., 2003). Desde la década pasada se han reportado un gran número de trabajos sobre los problemas involucrados en SLAM. Específicamente, se han propuesto diversos enfoques para construir mapas, como el topológico (Franz et al., 1998), el métrico (Movarec y Elfes, 1985), y enfoques híbridos que combinan los dos anteriores (Guivant et al., 2004; Kuipers et al., 2004; Folkesson y Christensen, 2004; Bosse et al., 2004; Zivkovic et al., 2005), y se han identificado aspectos críticos para lograr implementaciones prácticas y robustas de los algoritmos de SLAM, como la asociación de datos (Folkesson y Christensen, 2004), y la ambigüedad perceptual (Savelli y Kuipers, 2004).

Las siguientes secciones describen los diversos enfoques de mapeo, así como los desafíos relacionados con la asociación de datos y la ambigüedad perceptual.

II.2.1. Mapeo Topológico

Los enfoques topológicos representan el ambiente del robot como una gráfica de lugares conectados a través de arcos, sin depender de información métrica o geométrica, sino únicamente considerando las nociones de proximidad y orden. Los mapas topológicos se consideran representaciones compactas ya que sólo almacenan lugares interesantes y no el ambiente completo. Sin embargo, un conocido problema relacionado con este enfoque es el reconocimiento de lugares en ambientes complejos debido a la falta de información métrica que posibilite la discriminación entre dos lugares que parezcan similares. Así que en general, esta técnica funciona bien en ambientes estructurados y cerrados (tipo oficina), donde existen lugares distinguibles que se tratan como metas para el robot.

Un estudio en el que se propuso representar el ambiente partiendo únicamente de conocimiento topológico y no métrico, fue el de Franz y colaboradores (1998). Ellos plantearon un sistema basado puramente en visión para construir el mapa topológico de un ambiente abierto. La gráfica resultante sólo

almacena las vistas que permiten la descripción de las rutas relevantes en el ambiente, y la adyacencia entre las vistas, no el movimiento que conduce de una a otra. La única entrada sensorial al robot eran fotografías tomadas del ambiente a través de su cámara local. El problema reflejado en los resultados de este trabajo consistió en que sólo las áreas del ambiente suficientemente distinguibles pudieron ser mapeadas de manera confiable en la gráfica, dejando fuera otras áreas o vistas que también se consideraban representativas.

II.2.2. Mapeo Métrico

Desde la perspectiva métrica, el ambiente del robot es definido a través de un sistema coordinado global en el que el mapeo tiene lugar. Típicamente, el mapa es una matriz en la que cada celda representa cierta cantidad de espacio en el mundo real, lo cual hace que este enfoque funcione bien en ambientes acotados, pues se requiere de gran cantidad de memoria para representar grandes áreas. Además, el robot debe reconocer el lugar en el que se encuentra dentro del mapa para poder utilizarlo en la navegación y agregarle nueva información.

Los algoritmos que permiten construir un mapa métrico del ambiente comparten la característica de ser probabilísticos, lo cual se deriva del hecho de que el mapeo se caracteriza por la incertidumbre y por el ruido de los sensores. Las técnicas probabilísticas ofrecen un mecanismo para fusionar en el mapa los datos provenientes de diversos sensores, y esta fusión hace que el ruido disminuya.

El esquema dominante para integrar los datos provenientes de los sensores y los movimientos que ejecute el robot en diversos tiempos es el filtro de Bayes, el cual permite estimar el mapa y la posición del robot en el tiempo t a partir de las medidas de los sensores y de los movimientos del robot hasta el tiempo t .

Entre los algoritmos clásicos de mapeo métrico se encuentran los filtros de Kalman (Kalman, 1960), que son filtros de Bayes que representan las estimaciones a través de distribuciones Gaussianas, los algoritmos de Expectativa – Maximización conocidos como EM (Dempster et al., 1977), y las retículas de ocupación (Movarec y Elfes, 1985).

En el enfoque del filtro de Kalman (1960), los mapas son representados comúnmente a través de las coordenadas cartesianas de un conjunto de características del ambiente, las cuales pueden ser landmarks, objetos distintivos o formas. Así, el estado del ambiente comprende la posición del robot modelada por tres variables: sus coordenadas cartesianas en el plano, la dirección de su cabeza, y las coordenadas cartesianas del conjunto de características del ambiente.

Una ventaja del enfoque del filtro de Kalman es que la estimación del mapa se realiza en línea mientras el robot explora el ambiente. Además, converge con probabilidad uno al verdadero mapa y posición del robot, y es capaz de mapear ambientes cíclicos y ambientes dinámicos. Sin embargo, presenta problemas al tratar de identificar si dos lecturas de los sensores que parecen similares corresponden o no a la misma característica del ambiente. Por esta razón, este enfoque requiere que las características del ambiente sean suficientemente distintas, ya sea por la lectura de los sensores o por su localización, para que puedan ser identificadas confiablemente. Los mapas que resultan contienen las localidades de una pequeña cantidad de tipos de landmarks, careciendo de descripciones geométricas detalladas del ambiente.

Una alternativa al paradigma de los filtros de Kalman es la familia de algoritmos EM. EM genera mapas consistentes de ambientes cíclicos de gran escala aún si todas las características se parecen y no son distinguibles perceptualmente. Sin embargo, EM no involucra la noción de incertidumbre, sino que construye varios mapas para encontrar el que más se ajuste al ambiente. Para ello, EM tiene que procesar los datos muchas veces, por lo que no puede generar mapas de manera incremental como lo hacen los filtros de Kalman.

EM explota el hecho de que determinar un mapa cuando se conoce la trayectoria del robot es relativamente simple, como lo es determinar una estimación probabilística de la localización del robot en un

mapa dado. Esto lo logra iterando entre dos pasos: un paso de expectativa, en el que la posición del robot se calcula para un mapa dado, y un paso de maximización, en el que calcula el mapa que más se ajusta a la expectativa de la posición del robot. El resultado de la iteración de los pasos es una serie de mapas exactos, donde el primero es un mapa vacío. La principal desventaja de EM es que se ejecuta fuera de línea.

Por otro lado, el algoritmo que goza de mayor popularidad es el de retículas de ocupación. El problema central consiste en generar un mapa consistente a partir del ruido y de datos incompletos provenientes de los sensores. Aún cuando se conozca la posición del robot, no es fácil determinar si un espacio del ambiente está ocupado o no debido a ambigüedades en la lectura de los sensores.

Como el nombre del algoritmo lo sugiere, los mapas son retículas bidimensionales que representan un área fija en el sistema absoluto de coordenadas. Las retículas tienen alta resolución, en el orden de 5-10 cm por celda. Se utilizan filtros de Bayes para predecir la ocupación de cada celda. La actualización de los valores individuales de las celdas es incremental conforme llegan nuevos datos de los sensores. Al inicio, el valor de las celdas es 0.5, pues se desconoce si están o no ocupadas.

Las retículas de ocupación ofrecen representaciones muy completas del ambiente y útiles para evadir obstáculos y planear trayectorias, pero no ofrecen estimaciones consistentes a nivel global cuando trabajan en ambientes grandes (Guivant et al., 2004).

II.2.3. Mapeo Híbrido

En los últimos años, muchos estudios han presentado implementaciones de mapas a partir de la combinación de diferentes enfoques.

En 1991, Kuipers y Byun (1991), por ejemplo, propusieron una representación del ambiente llamada la jerarquía semántica espacial híbrida. En este modelo jerárquico, el elemento central es una red topológica con nodos que corresponden a lugares distintivos y arcos que corresponden a trayectorias de viaje. Un lugar que corresponda a un nodo debe ser distinguible dentro de su vecindad inmediata por algún criterio definible en términos de la entrada sensorial. Así, se determinan medidas de distinción a partir de un conjunto de características sensoriales, dentro del que alguna propiedad pueda ser maximizada en un lugar distintivo. Los autores consideran la firma de un lugar distintivo como el conjunto de características sensoriales, las medidas de distinción y los valores de las características que se maximizan en el lugar. Las trayectorias de viaje son definidas por estrategias de control que describen cómo el robot puede seguir la liga que conecta dos lugares distintivos.

Cada componente del mapa topológico es descrito en el nivel de control del modelo jerárquico en términos de estrategias de control y medidas de distinción. Además, cada componente tiene también información geométrica local, la cual constituye el nivel métrico del modelo. Esta información puede incluir la distancia y la dirección hacia objetos cercanos, la forma de tales objetos, la longitud y el ancho de la trayectoria, y el cambio en su orientación.

La estrategia de exploración que sigue el robot consiste de ciclos que incluyen los siguientes pasos: (i) a partir de un lugar, el robot se mueve en dirección abierta; (ii) selecciona una estrategia de control y sigue la trayectoria; (iii) detecta una vecindad, selecciona una medida de distinción y busca el lugar en el que se maximiza; (iv) alcanza el lugar distintivo que corresponda y el ciclo se repite. Así, la construcción del mapa topológico se efectúa mientras el robot se mueve entre las transiciones de la gráfica.

La posición del robot es descrita en el nivel topológico y en el métrico. En el topológico se describe a través de un lugar distintivo o por una pareja que representa una trayectoria y una dirección. En el métrico, cuando el robot está en un lugar distintivo, se describe su posición como la información sensorial actual y su orientación; cuando está en una trayectoria, se describe como el lugar del que viene, la distancia que ha recorrido y su orientación actual.

Recientemente, Kuipers y colaboradores (2004) adaptaron la jerarquía semántica espacial híbrida, construyendo mapas perceptuales como retículas de ocupación en regiones locales del ambiente, y utilizando el mapa topológico para describir la estructura del espacio a gran escala.

Thrun y colaboradores (1998), por otro lado, también llevaron a cabo el mapeo integrando una representación basada en retícula de ocupación y una representación topológica. En este caso, Thrun entrenó a una red neuronal artificial del tipo *back-propagation* para mapear las mediciones del sonar a los valores de ocupación para cada celda (x, y) de la retícula. La entrada a la red consiste de las cuatro lecturas sensoriales más cercanas a (x, y) , además de dos valores que codifican (x, y) en coordenadas polares relativas al robot (el ángulo al primero de los cuatro sensores, y la distancia). La salida de la red es 1, si (x, y) está ocupada, y 0, de otro modo. Los ejemplos de entrenamiento fueron obtenidos al hacer que el robot (en un simulador) operara en un ambiente conocido, y registrando y etiquetando las lecturas de los sensores. Una vez entrenada, la red generaba valores $[0, 1]$ que podían ser interpretados como probabilidades de ocupación.

Thrun incluyó una segunda fuente de información de ocupación: un sistema de cámara estéreo que proveía parejas de imágenes grabadas simultáneamente desde diferentes puntos de vista espaciales, las cuales se utilizaban para obtener información de profundidad, estimando así la proximidad de los obstáculos y proyectándola sobre la retícula de ocupación. Las interpretaciones sensoriales eran integradas en el tiempo para construir un mapa consistente, utilizando el filtro de Bayes.

Sobre los mapas basados en retículas de ocupación, se construyeron mapas topológicos. Al respecto, el algoritmo observado fue el siguiente: sobre la retícula de ocupación se construye un diagrama de Voroni, el cual es la unión de los puntos que forman un camino equidistante a los dos puntos ocupados más cercanos llamados puntos base; se encuentran los puntos críticos del diagrama de Voroni, los cuales minimizan localmente la distancia entre el punto (x, y) y los dos puntos base; se obtienen líneas críticas, conectando cada punto crítico a sus puntos base; las líneas críticas dividen el espacio libre de la retícula en regiones, las cuales forman los nodos de una gráfica topológica, donde los arcos corresponden a las líneas críticas.

Considerando trabajos más recientes, en (Zivkovic et al., 2005) se propone un algoritmo para construir mapas topológicos a partir de imágenes, utilizando una representación basada en apariencia, en la que el ambiente se modela como una colección de imágenes omnidireccionales tomadas en muchos lugares. Primero se construye una gráfica que incluye todas las imágenes, usando un criterio de agrupación que considera la presencia de landmarks visuales y las restricciones impuestas por la geometría del ambiente. Posteriormente, se emplea otro criterio de agrupación, de tal forma que las imágenes tomadas en un área convexa (un cuarto, por ejemplo) sean incluidas en un mismo grupo. Este algoritmo fue propuesto con el afán de ser empleado en los enfoques jerárquicos que combinan mapas conceptuales de alto nivel, usualmente topológicos, con mapas locales, geoméricamente precisos, de bajo nivel.

En (Bosse et al., 2004) se propone construir una gráfica de sistemas coordenados. Los nodos representan sistemas locales a manera de mapas métricos que modelan a nivel local el ambiente, y capturan la posición del robot y la correspondiente incertidumbre con respecto al propio sistema. Los arcos representan la transformación entre sistemas adyacentes. Esta propuesta, llamada Atlas, pretende servir como un marco de referencia genérico en el que una variedad de técnicas pueden ser usadas como el módulo de mapeo local.

En (Folkesson y Christensen, 2004) se utilizan aspectos característicos del ambiente para construir un mapa topológico con información métrica asociada con los arcos. Se toman mediciones de un conjunto de aspectos relativas a la posición actual del robot. El robot se desplaza entre tales mediciones, por lo que se estima también el cambio en la posición. Inicialmente, todas las estimaciones de posiciones se almacenan en un vector de estado, y a cada medición se asocia un valor de energía. Se construye una gráfica formada de nodos de estado para almacenar el vector de estado, y nodos de energía para los valores de energía. Se tienen arcos entre los nodos de energía y los nodos de estado que se requieren para calcular esa energía. Los nodos de estado pueden ser de posición o de algún aspecto característico. Cada nodo de posición se conecta a los nodos de posición previa y siguiente a través de nodos de energía que representan la odometría

del robot entre ellas. Un nodo de posición también puede conectarse a nodos de aspectos característicos, si es que el aspecto en cuestión fue percibido desde aquella posición.

Finalmente, en (Guivant et al., 2004) se presenta una estructura métrica híbrida que combina mapas basados en aspectos característicos paramétricos, como líneas, cilindros y esquinas, con retículas de ocupación. El mapa global de aspectos característicos es dividido en un conjunto de regiones triangulares locales, las cuales se utilizan como referencia para una descripción multidimensional del ambiente. En esta estructura, las retículas de ocupación soportan la asociación de datos y facilitan la extracción e incorporación de nuevos landmarks conforme son identificados desde muchos puntos estratégicos.

II.2.4. Asociación de Datos y Ambigüedad Perceptual

La **asociación de datos** se relaciona con el problema de determinar si dos características observadas en dos momentos diferentes corresponden o no a uno y al mismo objeto o lugar en el mundo físico (Hähnel et al., 2003). Este problema también ha sido referido como identificación de landmarks, donde las observaciones de landmarks empatan con las almacenadas dentro del mapa o ayudan a definir nuevas (Frese, 2006).

La **ambigüedad perceptual**, por otro lado, ocurre al tratar de distinguir entre lugares del ambiente que proveen patrones visuales equivalentes (Kuipers et al., 2004). Éste es el problema de falsa asociación de un landmark, en el que diferentes landmarks son considerados como el mismo por error (Frese, 2006).

Tanto las asociaciones incorrectas entre nuevos datos sensoriales y lugares previamente mapeados, como las distinciones incorrectas entre lugares perceptualmente similares, pueden conducir a fallas catastróficas en los algoritmos de SLAM. Tales fallas se evidencian cuando el robot trata de regresar a una región previamente mapeada después de una larga excursión, un problema referido como **cierre de ciclos** (Savelli y Kuipers, 2004; Bosse et al., 2004).

Los métodos basados en apariencia y las técnicas de multi-hipótesis abordan el problema de la asociación de datos. Las firmas de apariencia y las métricas de similitud de imágenes relacionadas con el color, la forma y la textura han sido propuestas para reconocer lugares en enfoques topológicos de mapeo (Ulrich y Nourbakhsh, 2000), y recientemente, las medidas de apariencia han sido aplicadas también para detectar ciclos en enfoques métricos de SLAM (Newman et al., 2006). Las técnicas multi-hipótesis, por otro lado, manejan situaciones en las que puede resultar imposible determinar la identidad de un landmark al momento de observarla, y se requiere diferir la decisión acerca de la asociación de datos hasta que observaciones posteriores provean evidencia suficiente. Para ello, estas técnicas generan múltiples copias del mapa incierto cada vez que la identidad de un landmark es ambigua, lo cual agrega una importante carga computacional. Algunos ejemplos de estas técnicas incluyen el filtro de Kalman multi-hipótesis (Bar-Shalom y Fortmann, 1988), y los algoritmos basados en filtros de partículas como FastSLAM (Montemerlo y Thrun, 2003). En este último caso, tanto la incertidumbre discreta de la identificación de landmarks como la incertidumbre continua producida por el ruido en las mediciones son representadas por un conjunto de partículas, donde cada partícula corresponde a un mapa sin incertidumbre. En caso de que el mapa contenga varias ambigüedades independientes, el filtro requerirá incrementar exponencialmente la cantidad de partículas para cubrir todas las posibles combinaciones, lo cual implica el uso de un algoritmo de asociación de datos capaz de revisar decisiones pasadas de asociación de datos cuando nueva evidencia sugiere que han sido incorrectas (Hähnel et al., 2003).

II.2.5. Evolución de los Algoritmos de SLAM

De acuerdo con Frese (2006), las etapas iniciales de los algoritmos de SLAM pueden ser ubicadas dentro de la década que abarca la segunda mitad de los 80 y la primera de los 90. Durante este periodo, la

formulación matemática de SLAM estaba abierta y el problema de la incertidumbre no se reconocía completamente. Fue entonces cuando se propusieron los enfoques clásicos de mapeo métrico como el de retículas de ocupación (Movarec y Elfes, 1985).

El término SLAM no fue introducido sino hasta 1995 (Durrant-Whyte et al., 1995), habiendo transcurrido una década en la que el tiempo computacional de los algoritmos era la principal preocupación, pues restringía el uso de SLAM a ambientes pequeños que incluyeran menos de 100 landmarks. Actualmente se han propuesto muchos nuevos algoritmos más eficientes que pueden ser utilizados en ambientes medianos de 500 landmarks aproximadamente, y algunos muy rápidos que se emplean en ambientes grandes con más de 1000 landmarks, aunque éstos presentan algunas limitaciones referentes a la calidad del mapa estimado en ciertas situaciones (Frese, 2006).

Como se ve, la mayor parte de la investigación en SLAM versa sobre la propuesta de algoritmos computacionalmente más eficientes, de soluciones a los problemas de asociación de datos y ambigüedad perceptual, y de representaciones ambientales de distintos tipos. Sin embargo, pocos esfuerzos han sido documentados en relación al uso de tales representaciones espaciales para navegar directamente hacia localidades meta designadas. Aún más, no hemos encontrado reportes de algoritmos propuestos para manejar el proceso de desaprendizaje de localidades meta previamente aprendidas ante cambios en el ambiente. De aquí que uno de los propósitos del presente trabajo consista en abordar la inequidad entre mapeo y explotación del mapa que existe en la literatura sobre SLAM, así como la falta de investigación sobre desaprendizaje, tomando inspiración de tales habilidades en las ratas.

II.3. TRABAJO RELACIONADO

Los descubrimientos biológicos relacionados con el aprendizaje y la memoria espacial de animales como las ratas han sido considerados por muchos investigadores como atractivos para incorporar sus principios en modelos de cognición espacial y navegación aplicados a robots móviles. Esta sección introduce algunos de los esfuerzos más importantes documentados en los últimos años estableciendo comparaciones con el trabajo aquí propuesto.

II.3.1. El modelo de Burgess y colaboradores (1994)

En su forma original, el modelo de Burgess y colaboradores (1994) consiste de cinco capas. Información espacial del ambiente dispara células sensoriales, la actividad de éstas se propaga a través de las células de la corteza entorhinal hasta llegar a las células de lugar del hipocampo. La salida de las células de lugar converge después, a través de células intermediarias del subiculum, en una capa de salida que consiste en una pequeña población de células de meta hipotéticas. Posteriormente, cuando los autores implementan su modelo en un robot, eliminan la capa de células del subiculum.

La entrada al modelo es tomada, entonces, de conjuntos de células sensoriales, donde cada conjunto codifica la distancia hacia una señal externa específica y cada célula del conjunto genera su máxima respuesta ante una distancia particular.

El modelo utiliza el principio de codificación de fase en todos los niveles de procesamiento de la información espacial en la red neuronal. La codificación de fase se refiere a lo siguiente. El electroencefalograma (EEG) del hipocampo es la señal eléctrica más larga en el cerebro de la rata. Una forma del EEG, llamada ritmo theta, es una oscilación de 7 – 12 Hz. O'Keefe y Nadel (1978) sugirieron que en la

rata el ritmo theta coincide con los movimientos de desplazamiento de la misma; es decir, las células de lugar disparan en fase con el ritmo theta y esta relación de fase cambia sistemáticamente cuando la rata se mueve a través del campo de lugar de una célula de lugar específica. Las células de lugar que disparan en una fase tardía tienen campos de lugar centrados delante de la rata, mientras que aquellas que disparan en una fase temprana tienen campos de lugar centrados atrás de la rata en ambientes abiertos (Burgess y O'Keefe, 2002).

La transformación de la respuesta de las células sensoriales en campos de lugar bien localizados es realizada por las células de la corteza entorhinal. Cada una de estas células recibe conexiones de dos células sensoriales. La fase de disparo de una célula entorhinal es relacionada con el ángulo promedio de la dirección de la cabeza de la rata a las dos señales que proveen la entrada sensorial a la célula. Si el ángulo está delante de la rata, la célula dispara en una fase tardía del ritmo theta. Si el ángulo está atrás, la célula dispara en fase temprana. Lo anterior significa que las células de la corteza entorhinal que disparan tarde en el ciclo theta están respondiendo a la entrada sensorial de las parejas de señales cuyo centro se encuentra delante de la rata. A partir de aquí se construye una tendencia para la fase de disparo de las células de las demás capas del modelo, dado que la activación de las células de la corteza entorhinal alimenta hacia delante a la red neuronal.

La última capa de la red incluye cuatro células de meta, cada una representa una dirección aloéctrica (relativa al mundo, no a la rata) en el ambiente (norte, sur, este, oeste). Una célula de meta que representa una dirección específica recibe una fuerte entrada excitatoria cuando la rata se encuentra en el lugar de la meta y su cabeza está dirigida hacia tal dirección. En realidad, la localidad de una meta es representada en el modelo por un vector de cuatro células de meta; esto es, la suma del vector de dirección asociado a cada célula de meta ponderado por la razón de disparo de ésta es un vector que indica la dirección de la rata desde la meta. El modelo asume que las células de meta tienen entradas del sistema que calcula la dirección de la cabeza de la rata y el sistema que administra la recompensa, de tal forma que las conexiones entre las células de lugar y las células de meta pueden ser modificadas cuando la rata encuentra la meta y ve hacia la dirección correcta. Dicha modificación ocurre en una fase tardía del ciclo theta y enciende las conexiones que salen de las células de lugar activas en ese momento. Esto implica que la célula de meta asociada al norte recibe fuertes conexiones de las células de lugar cuyos campos de lugar están centrados hacia el norte de la meta (esto es, células de lugar activas en la fase tardía del ciclo theta cuando la rata está viendo al norte). Lo mismo ocurre para las células de meta asociadas a las otras direcciones.

Cuando el robot encuentra el lugar de la meta, una señal de refuerzo provoca un disparo de aprendizaje de Hebb (1949) en las conexiones que van de las células de lugar a las células de meta. Las subsecuentes tasas de disparo de estas células proveen una estimación continua de la dirección y proximidad del lugar de la meta, posibilitando la navegación.

Como se ha dicho, la información métrica relacionada con las distancias hacia señales visuales constituye la entrada exclusiva y directa al sistema. En contraste, las celdas de lugar en nuestro modelo codifican la integración de información visual y kinestésica, y adicionalmente, las propiedades de las señales visuales son interpretadas por medio de neuronas sensibles a patrones específicos de información de landmarks. Finalmente, nuestro modelo agrega la construcción explícita de un mapa topológico de lugares y de sus relaciones métricas, no limitándose a la generación de campos de lugar aislados.

II.3.2. El modelo de Redish y Touretzky (1997)

En el modelo de Redish y Touretzky (1997), al igual que en el nuestro, la representación de los lugares integra información vestibular y visual. La entrada visual consiste del tipo de cada landmark, así como su distancia y orientación con respecto al animal. La dirección de la cabeza de la rata se actualiza mediante claves vestibulares y comandos motrices ejecutados por el animal. El modelo asume que el animal mantiene una estimación de su posición de acuerdo con un sistema coordinado interno. Este valor es actualizado por

un módulo integrador de trayectoria mientras el animal se mueve considerando la entrada vestibular y de locomoción.

La codificación de lugares en el modelo de los autores utiliza unidades de lugar para mantener la asociación entre la percepción local del animal y las coordenadas definidas por el integrador de trayectoria. Como en nuestro modelo, cuando no existen señales visuales en el ambiente, las unidades de lugar son disparadas únicamente por el integrador de trayectoria.

En el modelo de Redish y Touretzky, las unidades de lugar se crean dinámicamente mientras el animal simulado explora el ambiente, a diferencia del nuestro, en donde la cantidad de células de lugar se encuentra predeterminada y no depende del ambiente que la rata explore.

Para determinar si el animal reconoce las señales percibidas en su localidad actual, los autores utilizan una medida de consistencia que resulta alta cuando la vista local actual es compatible con la vista local recordada por las coordenadas actuales del integrador de trayectoria, o cuando estas coordenadas empatan con las que se encuentran asociadas con la vista local actual. Este proceso de reconocimiento de lugares es similar al que lleva a cabo nuestro sistema sobre la representación espacial. En este sentido también contrastan los modelos, pues el nuestro no define la representación espacial a manera de campos de lugar aislados, sino que construye un mapa topológico-métrico holístico o global considerando los patrones de actividad derivados de la población completa de unidades de lugar para definir lugares distintivos y sus relaciones.

Información relativa a la motivación del animal es utilizada en el modelo de Redish y Touretzky para planear una ruta hacia una meta dada, proceso que los autores relacionan con el estriado de la rata recibiendo entradas espaciales del hipocampo. En nuestro sistema, la motivación y el aprendizaje juegan un rol fundamental también, y adicionalmente, modelamos la habilidad de desaprendizaje de la rata cuando las localidades meta cambian, y sugerimos la influencia del estriado sobre el hipocampo a través del giro dentado para permitir al animal explotar las expectativas de futura recompensa durante tareas espaciales reforzadas. Esta sugerencia se explica en el Capítulo III de la disertación.

II.3.3. El modelo de Guazzelli y colaboradores (1998)

El estudio llevado a cabo por Guazzelli y colaboradores (1998) propone un modelo referido como TAM-WG, el cual provee dos sistemas de navegación (O'Keefe y Nadel, 1978): uno para llegar directamente hacia una meta perceptible desde la localización del animal (*taxon*), y otro para relacionar la posición de la meta con la localización de landmarks externos para poder reconocer tal posición objetivo dentro del mapa cognitivo y navegar hacia ella (*locale*). La representación espacial provista por el modelo para posibilitar este último sistema de navegación combina información kinestésica y visual. El modelo fue validado mediante simulaciones en ambientes virtuales de algunas tareas espaciales clásicas que han sido implementadas con ratas.

Como se describe en el Capítulo III de la disertación durante la presentación del modelo propuesto, nuestro trabajo se encuentra parcialmente inspirado en el de Guazzelli y colaboradores. Una de las principales extensiones propuestas a este sistema consiste en un proceso de explotación del mapa para posibilitar la navegación orientada a metas en el robot móvil. El modelo original asigna a la rata simulada la capacidad de aprender localidades meta desde un punto de partida fijo dentro de laberintos que incluyen únicamente un punto de decisión. No obstante, el animal es incapaz de lograr dos aspectos fundamentales: (i) encontrar el objetivo en laberintos más complejos que incluyan dos o más localidades en las que la rata requiera decidir su próxima orientación, y (ii) alcanzar la meta a partir de localidades iniciales arbitrarias en el laberinto aprendido.

Adicionalmente, el modelo original fue extendido para incluir un proceso de adaptación del mapa que permite representar en línea cambios en la configuración física del ambiente percibidos por el robot, como el intercambio de posición de algunos landmarks o la eliminación de uno de ellos.

Aún cuando también ha sido nuestro interés probar al robot móvil en tareas espaciales clásicas llevadas a cabo con ratas, hemos validado nuestra arquitectura robótica de manera adicional diseñando e implementando nuevos experimentos con ratas que produjeron datos conductuales comparables con los resultados derivados del robot.

Finalmente, a diferencia de Guazzelli y colaboradores, el modelo propuesto sugiere, como ya se ha mencionado, la influencia del estriado sobre el hipocampo a través del giro dentado en el cerebro de la rata.

II.3.4. El modelo de Gaussier y colaboradores (2002)

En el modelo de Gaussier y colaboradores (2002), las células de lugar o células de vista son aprendidas en la corteza perirhinal y en la corteza entorhinal, y el rol del hipocampo no consiste fundamentalmente en la construcción de mapas cognitivos o en la navegación, sino que es utilizado para aprender, almacenar y predecir transiciones entre diversos estados.

Cada célula de lugar codifica un lugar del ambiente, y la regla de Hebb permite el aprendizaje de la relación de tiempo (y por tanto de la relación topológica) entre dos lugares adyacentes. El modelo produce una gráfica de las relaciones espaciales entre los lugares.

La navegación implica el aprendizaje del vínculo entre el reconocimiento de un lugar dado y el alcance de la meta. Para ello, el modelo asocia el reconocimiento de ese lugar con la activación de una neurona motivacional. La conexión entre tal neurona y la célula de lugar que representa la situación reconocida es reforzada utilizando la regla de Hebb, y durante el proceso de navegación, la actividad motivacional se propaga hacia atrás dentro del mapa. Posteriormente, el sistema trata de reconocer la localización actual del robot, y selecciona como submeta el nodo más activo del mapa cognitivo directamente ligado con su posición actual. El sistema alcanza ese lugar y el proceso se repite (selección de la siguiente submeta) hasta alcanzar la meta más cercana a su posición inicial.

Para posibilitar la selección de las acciones a ejecutar para alcanzar la meta, los autores construyen una representación de la transición entre dos situaciones o lugares, la cual es asociada con el movimiento para ir de un lugar a otro. Posteriormente, las conexiones entre las representaciones son aprendidas para crear una gráfica de las relaciones topológicas entre las transiciones. Cuando el robot alcanza la meta, aprende la asociación entre el reconocimiento de la última transición y la satisfacción de la motivación. Durante la navegación, la motivación se propaga hacia atrás hasta alcanzar el estado actual, permitiendo la activación de los nodos de la gráfica e indicando el movimiento que se debe ejecutar para llegar a la meta.

Los autores proponen que el mapa cognitivo podría ser elaborado en la corteza prefrontal, pero que las representaciones internas de las transiciones podrían ser codificadas en las áreas CA3-CA1 del hipocampo.

El modelo fue probado con un robot móvil en un ambiente abierto. En una primera fase, el robot explora el ambiente de manera aleatoria, y elabora progresivamente el mapa cognitivo creando un nuevo nodo por cada nueva transición y ligando nodos conforme son encontrados consecutivamente. El movimiento asociado con cada transición se obtiene después de la integración de los movimientos que realiza el robot desde el lugar previamente activo hacia el siguiente. La integración se reinicia cuando el robot descubre un nuevo lugar o llega a un lugar conocido. El experimento consistió únicamente en el aprendizaje de un camino muy simple formado por cinco transiciones. Después de la corta exploración, el robot es colocado nuevamente en la posición inicial y se activa la neurona motivacional. Debido al aprendizaje, la motivación se propaga hacia atrás en el mapa cognitivo. Al mismo tiempo, el robot reconoce su posición actual (la célula de lugar más activa) a partir del panorama visual. Este reconocimiento es usado para predecir todas las posibles

transiciones desde la posición actual. Cuando varias transiciones son factibles, la propagación hacia atrás de la meta en el mapa permite influenciar o resolver la competencia entre las transiciones predichas. La mejor transición, de acuerdo con el sistema de planeación, es seleccionada.

Aún cuando el experimento resultó exitoso, los autores reconocen que con su modelo no es posible distinguir entre lugares visualmente similares que aparecen en corredores largos o en condiciones de oscuridad, lo cual no ocurre en nuestro modelo debido al uso de información kinestésica además de información visual derivada de landmarks externos como entrada al sistema. Otros aspectos que diferencian nuestro modelo del de Gaussier y colaboradores incluyen el rol de las células de lugar del hipocampo codificando transiciones entre estados en vez de lugares, y el proceso de reconocimiento de lugares que lleva a cabo la corteza entorhinal en vez del hipocampo. Como es nuestro modelo, los autores construyen una representación topológica del espacio, aunque los nodos de este mapa no corresponden a lugares sino a transiciones entre estados, e implementan un proceso de explotación del mapa para determinar secuencias de transiciones entre estados que conducen a una localidad meta. Sin embargo, no modelan la motivación del animal en relación a su necesidad de comer ni la predicción de expectativas de recompensa.

II.3.5. El modelo de Arleo y colaboradores (2004)

En el modelo de Arleo y colaboradores (2004), estímulos externos (datos visuales) son interpretados por células especializadas llamadas de fotografía para caracterizar las regiones del ambiente mediante patrones sensoriales distintos, lo cual resulta en una representación espacial alotética, y estímulos internos (kinestésicos y vestibulares) son integrados en el tiempo para proveer una representación idiotética. Ambas representaciones, alotética e idiotética, se combinan para formar una representación espacial en el hipocampo que consiste de un gran número de campos de lugar traslapados.

Cada vez que el robot se encuentra en un nuevo lugar del ambiente, todas las células de fotografía activas simultáneamente se conectan a una nueva célula de lugar del hipocampo. A la nueva sinapsis se le asigna un peso aleatorio en $[0, 1]$, y posteriormente su eficacia cambia de acuerdo a la regla de aprendizaje de Hebb. Cuando el robot visita un lugar, verifica primero si existen células de lugar que lo codifiquen. Conexiones de las células de fotografía a nuevas células de lugar son creadas sólo si la cantidad de neuronas de lugar activas no excede un umbral específico.

Por otro lado, el modelo incluye células de integración de trayectoria, cuya actividad cambia de acuerdo con los estímulos kinestésicos y con la orientación del robot, y es independiente del ambiente.

Para asociar la actividad de las células de lugar con una localización espacial, los autores aplican el esquema de codificación del vector de población, y promedian la actividad de la población neuronal para obtener la localización codificada.

Finalmente, para posibilitar la navegación, el modelo incorpora una población de células de acción cuya actividad provee comandos de dirección de movimiento. Para cada tipo de meta (agua o comida, por ejemplo), cuatro neuronas de acción (codificando las acciones alocéntricas norte, sur, este y oeste) son disparadas por la capa de células de lugar. Las sinapsis de las células de lugar a las neuronas de acción son modificadas para aprender una función que mapea localidades espaciales continuas en acciones motrices. Se utiliza aprendizaje por refuerzo para encontrar dicha función.

En síntesis, los componentes principales de la arquitectura neuronal propuesta por Arleo y colaboradores resultan similares a los encontrados en nuestro modelo: la combinación de información alotética (visual) y señales idiotéticas (integración de trayectoria) al nivel de la representación hipocámpal, el uso de aprendizaje de Hebb para correlacionar tales entradas, el mapeo de la actividad derivada de la población de células de lugar en localidades espaciales, y la aplicación de aprendizaje por refuerzo para soportar la navegación orientada a metas. Sin embargo, las siguientes diferencias existen. Para mapear la actividad grupal de las

células de lugar en localidades espaciales, nuestro trabajo reemplaza el esquema de codificación del vector de población por el uso de información derivada de las posibles rotaciones que el animal es capaz de ejecutar estando en un lugar específico orientado hacia una dirección particular. Por otro lado, agregamos la construcción explícita de un mapa topológico de lugares y de sus relaciones métricas. Al respecto, tanto el modelo de Arleo y colaboradores como el de Burgess y colaboradores (1994) utilizan la activación de las células de lugar para controlar directamente las acciones motrices a ejecutar por el robot a través de una capa de cuatro células de acción que codifica las acciones alocéntricas norte, sur, este y oeste. Finalmente, el modelo de Arleo no está ligado en absoluto con experimentos reales practicados con ratas, y aunque utiliza una técnica estándar de aprendizaje por refuerzo, no incluye el aspecto motivacional que caracteriza el comportamiento de los animales, y tampoco predice y adapta expectativas de recompensa durante la exploración del ambiente para ser explotadas durante la navegación orientada a metas.

II.3.6. El modelo de Milford y colaboradores (2006)

El trabajo realizado por Milford y colaboradores (2006) se interesa por la efectividad de los modelos del hipocampo de la rata en aplicaciones con robots móviles explorando ambientes grandes que incluyen señales visuales naturales. Nuestro interés, en cambio, consiste en asignar a robots móviles capacidades de cognición espacial similares a las encontradas en roedores, con el afán de producir comportamientos comparables y eventualmente proveer retroalimentación valiosa a neurocientíficos experimentalistas.

A pesar de la diferencia mencionada en cuanto al propósito de la investigación, nuestro modelo coincide con el de Milford y colaboradores en algunos aspectos relacionados con el mapeo y la adaptación del mapa, y contrasta con él en la navegación orientada a metas. Específicamente, el modelo de Milford y colaboradores construye un mapa topológico de experiencias donde cada experiencia representa una fotografía de la actividad registrada a partir de células de posición y de células de vista local en un tiempo dado. Las células de posición codifican la localización y orientación físicas del robot, mientras que las células de vista local codifican información visual y forman parte de la entrada a las células de posición. Una escena visual familiar activa una célula de vista local, que a su vez inyecta actividad en las células de posición asociadas con ella. Las transiciones entre las experiencias almacenadas en el mapa están asociadas con información referente al movimiento físico del robot cuando pasó de una experiencia a otra; en particular, la magnitud de la rotación y la duración de la transición. En nuestro caso, los nodos almacenados en el mapa topológico que el modelo construye también representan asociaciones entre patrones de información visual y señales de integración de trayectoria, y la actividad registrada en la población completa de células de lugar, mientras que las transiciones entre nodos están asociadas con información métrica derivada de la locomoción de la rata referente a su dirección alocéntrica al pasar de un nodo a otro y a la cantidad de pasos que ejecutó para hacerlo.

El mapa de experiencias de los autores puede ser adaptado ante cambios físicos del ambiente, lo cual implica la eliminación/creación de experiencias y la actualización de transiciones entre experiencias. Como se detallará más adelante en el Capítulo V de la disertación, nuestro sistema también adapta en línea el mapa construido para representar cambios en la configuración física de landmarks.

En cuanto al aspecto de navegación, la información temporal almacenada en el mapa de experiencias de Milford y colaboradores se emplea para encontrar la ruta más rápida que conduce a la meta, y posteriormente, la información espacial y conductual almacenada en las transiciones entre experiencias se utiliza para navegar hacia el objetivo. En contraste, como se ha mencionado, nuestro modelo considera el estado motivacional de la rata en una arquitectura de aprendizaje por refuerzo que permite al animal aprender y desaprender localidades de recompensa, y que soporta la navegación hacia la meta desde cualquier punto de partida dentro del ambiente.

III. MODELO DE COGNICIÓN ESPACIAL Y NAVEGACIÓN EN RATAS

El modelo propuesto de cognición espacial y navegación en ratas comprende diversos módulos funcionales ilustrados en la Figura III.1, los cuales capturan algunas propiedades de las estructuras del cerebro de la rata involucradas en el aprendizaje y la memoria. Este capítulo provee el contexto biológico que subyace al modelo computacional, así como la descripción detallada de sus módulos.

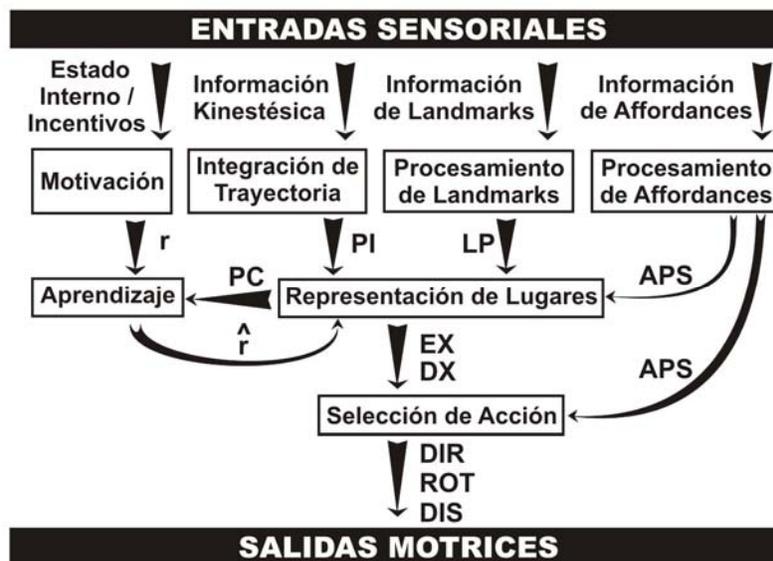


Figura III.1. Los módulos del modelo de cognición espacial y navegación en ratas, y su interacción. Glosario: r – recompensa primaria o inmediata; PI – patrón de información kinestésica; LP – patrón de información de landmarks; APS – esquema perceptual de affordances; PC – patrón de información de lugar; \hat{r} – señal de refuerzo efectivo; EX – expectativas de máxima recompensa y sus direcciones correspondientes (DX); DIR – próxima dirección de la rata; ROT – rotación a ejecutar por la rata; DIS – próximo desplazamiento del animal.

III.1. CONTEXTO BIOLÓGICO

La Figura III.2 ilustra el contexto biológico del modelo propuesto.

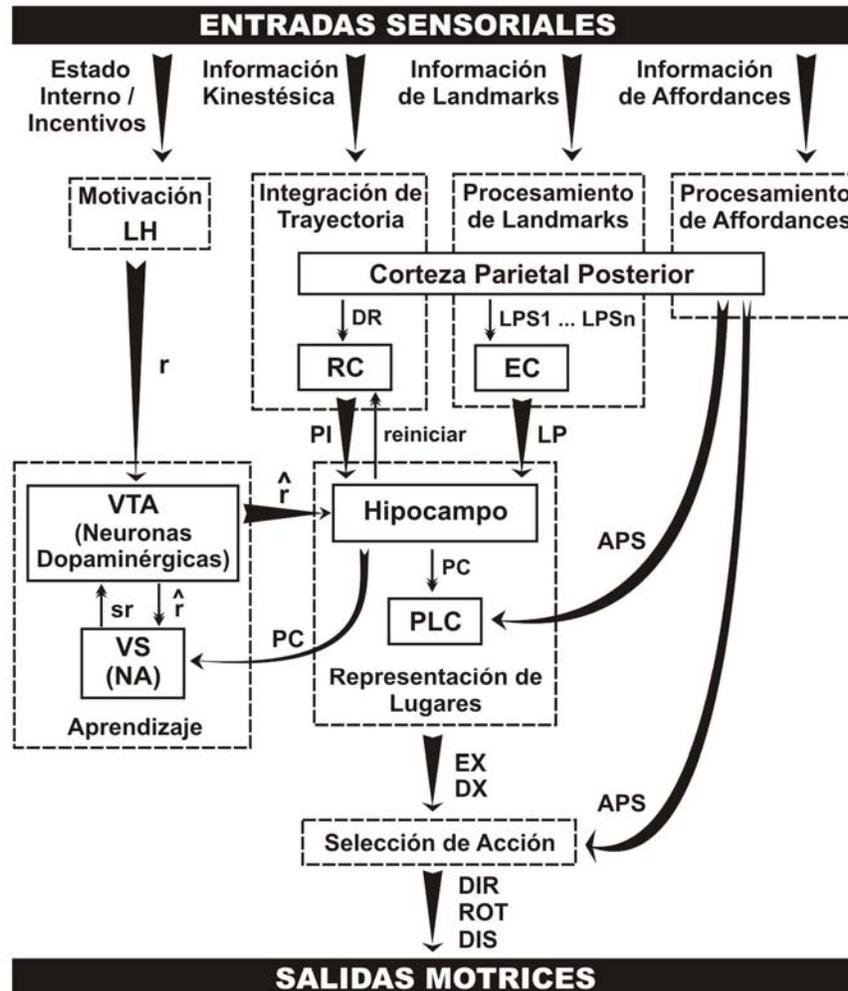


Figura III.2. Contexto biológico que subyace al modelo computacional de cognición espacial y navegación en ratas. Glosario: LH – hipotálamo lateral; RC – corteza retrosplenial; EC – corteza entorhinal; VTA – área ventral tegmental; VS – estriado ventral; NA – núcleo accumbens; PLC – corteza prelímbica. Entradas/Salidas: r – refuerzo primario; sr – refuerzo secundario; \hat{r} – refuerzo efectivo; DR – esquema perceptual de remapeo dinámico; LPS – esquema perceptual de un landmark; APS – esquema perceptual de affordances; PI – patrón de información kinestésica; LP – patrón de información de landmarks; PC – patrón de información de lugar; EX – expectativas de máxima recompensa y sus direcciones correspondientes (DX); DIR – próxima dirección de la rata; ROT – rotación a ejecutar por la rata; DIS – próximo desplazamiento del animal.

Como se mencionó en el Capítulo II, el hipotálamo es considerado como el área principal donde la información relacionada con el estado motivacional interno de la rata se combina con la presencia de incentivos (Risold et al., 1997). Específicamente, la búsqueda y la ingestión de comida son actividades que se asumen controladas por el hipotálamo lateral (Kelley, 2004). Por lo tanto, el módulo de motivación del modelo está relacionado funcionalmente con esta región del cerebro del animal, el cual calcula su valor de hambre y produce la recompensa primaria o inmediata que la rata obtiene ante la presencia de comida (r).

La corteza parietal posterior (PPC) es considerada como una estructura sensorial que recibe información multimodal de tipo kinestésico, visual y relativa a rotaciones del animal. El procesamiento que PPC ejecuta sobre cada tipo de información se describe enseguida.

Se ha sugerido que PPC forma parte de la red neuronal involucrada en el proceso de integración de trayectoria (Parron y Save, 2004), donde la corteza retrosplenial (RC) también está implicada (Cooper y Mizumori, 1999; Cho y Sharp, 2001). Considerando lo anterior, el modelo atribuye a PPC la representación de la posición actualizada del punto de partida de la rata en relación a su posición actual cada vez que el animal se

mueve, utilizando un esquema perceptual de remapeo dinámico (*DR*), y a RC, la generación y detección de patrones de información kinestésica (*PI*), tarea que lleva a cabo una de las capas del integrador de trayectoria del modelo. El hipocampo reinicia la posición del punto de partida del animal en *DR* cuando resulta necesario; por ejemplo, al inicio de un ensayo en un experimento dado.

Se asume que la corteza entorhinal (*EC*) está involucrada en el procesamiento de landmarks. De esta forma, *EC* recibe de PPC información espacial acerca de landmarks perceptibles (Redish, 1997), la cual consiste de la distancia y orientación relativa de cada landmark, ambas codificadas dentro de un esquema perceptual de landmark (*LPS*). Posteriormente, *EC* genera y detecta patrones de información de landmarks, y los integra en un solo patrón para representar la vista egocéntrica desde la posición del animal (*LP*).

Se ha demostrado que antes de que la rata ejecute cualquier movimiento, aproximadamente la mitad de las neuronas encontradas en PPC exhiben actividad relacionada con el movimiento distinguiendo entre modos básicos de locomoción como rotaciones a la izquierda, a la derecha o desplazamientos hacia delante (McNaughton et al., 1994(2)). Por lo tanto, el modelo atribuye a PPC la generación del esquema perceptual de affordances (*APS*), el cual codifica los posibles giros que la rata puede ejecutar en un momento dado estando en una localidad específica orientada hacia una dirección particular.

Por otro lado, el módulo de representación de lugares del modelo incluye una capa de células de lugar referida como PCL, y una capa que genera la gráfica del ambiente llamada WGL. El hipocampo recibe información kinestésica y visual procedente de RC y EC respectivamente. La actividad de las células de lugar resulta de la integración de ambas fuentes de información. Los campos de lugar traslapados en la colección de neuronas en PCL son asociados con un área física dentro del ambiente de la rata. Tal área es identificada de manera direccional por el patrón de actividad de la colección de neuronas (*PC*), y su extensión queda determinada por cambios de affordances detectados por el animal durante el proceso de exploración del ambiente. Conviene aclarar que la versión actual del modelo no distingue entre la dinámica de grupo de las regiones hipocampales CA3 y CA1 (Guzowski et al., 2004).

Las asociaciones entre campos de lugar traslapados y áreas físicas en el ambiente son representadas por la capa WGL del modelo a través de un mapa espacial holístico o global. Adicionalmente al proceso de mapeo, WGL ejecuta un proceso de reconocimiento de lugares. Hemos analizado dos hipótesis relacionadas con la situación de WGL dentro del cerebro de la rata: (i) el hipocampo, pues de acuerdo con (Hollup et al., 2001(1)), lesiones hipocampales provocan un severo y selectivo déficit en la identificación de una localidad, sugiriendo que el hipocampo puede resultar esencial para el reconocimiento de lugares durante la navegación espacial; y (ii) la corteza prelímbica, una subregión de la corteza frontal de la rata que está involucrada no sólo en la memoria de trabajo, sino también en gran cantidad de procesos requeridos para resolver problemas cognitivos complejos (Granon y Poucet, 2000), y en el control de comportamientos orientados a metas (Grace et al., 2007). Dado que el modelo depende de la explotación del mapa espacial mantenido por WGL para posibilitar la navegación dirigida por metas, hemos preferido asumir que la funcionalidad de WGL puede ser correspondida con la de la corteza prelímbica.

Se sabe que la información de recompensas es procesada en los ganglios basales del cerebro por medio de sus neuronas dopaminérgicas, las cuales responden ante recompensas primarias y secundarias, y sus respuestas reflejan "errores" en la predicción de recompensas que constituyen señales de enseñanza utilizables en el aprendizaje por refuerzo. Por otro lado, las neuronas del estriado ventral (núcleo accumbens) se activan cuando el animal espera recompensas predecibles, y adapta dicha actividad de expectativas a nuevas situaciones de recompensa (Schultz et. al, 1998). Houk y colaboradores (1995) propusieron que el estriado implementa una arquitectura Actor-Crítico (Barto, 1995), donde un Crítico Adaptivo predice el valor de recompensa de cualquier lugar (*PC*) en el ambiente y produce una señal de error (\hat{r}). Una cantidad específica de unidades Actor es incluida en esta arquitectura de aprendizaje para representar las posibles acciones a ejecutar por la rata. En el modelo, las unidades Actor corresponden a las posibles orientaciones que la rata pueda adoptar estando en cualquier lugar dado. Las expectativas de recompensa asociadas a tales orientaciones son adaptadas a través de \hat{r} . Recientemente, se ha sugerido que las ratas con lesiones en el giro dentado (*DG*) del hipocampo están severamente impedidas para resolver tareas reforzadas que

implican la memoria de trabajo, y que el desempeño que muestran durante este tipo de tareas está fuertemente correlacionado con la densidad de las células en DG pero no con la densidad de las células en las áreas CA1 y CA3 del hipocampo (Hernandez-Rabaza, et al., 2007). Entonces, considerando que dichas ratas se encontraron impedidas específicamente en su habilidad de actualizar información espacial esencial para guiar comportamientos orientados a metas, el modelo asume que las unidades Actor pueden localizarse en DG. De esta forma, el modelo propuesto sugiere que el estriado debiera influir al hipocampo a través de DG enviando las expectativas de futura recompensa correspondientes a las acciones del animal para ser aprovechadas por DG durante el desempeño apropiado del animal durante la ejecución de tareas espaciales reforzadas.

Finalmente, el módulo de selección de acción del modelo determina las salidas motrices del mismo, las cuales consisten de la próxima dirección de la cabeza de la rata, la rotación requerida para orientarse hacia tal dirección, y el desplazamiento a ejecutar.

III.2. MOTIVACIÓN

En el modelo, la motivación del animal está relacionada con su necesidad de comer: el instinto de hambre. De acuerdo con Arbib y Lieblch (1977), los instintos pueden ser deseables o no deseables. El módulo de motivación del modelo calcula el valor del hambre D en el instante de tiempo $t+1$ utilizando la ecuación (1), y la recompensa inmediata o refuerzo primario r que el animal obtiene ante la presencia de comida a partir de la ecuación (2). La recompensa depende del estado motivacional actual de la rata. Si está extremadamente hambrienta, la presencia de comida puede ser más estimulante que en caso de no estarlo.

$$D(t+1) = D(t) + \alpha_d |d_{\max} - D(t)| - a |D(t)| + b |d_{\max} - D(t)| \quad (1)$$

$$r(t) = D(t) / d_{\max} \quad (2)$$

Un instinto deseable se incrementa espontáneamente en cada instante de tiempo hasta alcanzar d_{\max} (valor constante), mientras que los instintos no deseables se reducen hasta llegar a 0, ambos según un factor α_d (valor constante) intrínseco al animal. Un incremento adicional ocurre si un incentivo b se presenta, como la percepción o el olor de la comida, y una reducción a se registra como consecuencia de la ingestión de comida.

III.3. PROCESAMIENTO DE AFFORDANCES

La noción de affordances de movimiento, adoptada de Gibson (1966), representa todas las posibles acciones motrices que la rata puede ejecutar y que determina a partir de la sensación inmediata del ambiente; por ejemplo, la percepción de un corredor le indica que puede desplazarse hacia delante, la sensación de brazos a partir de una localidad en un laberinto le indican que puede girar. En el modelo, las affordances de movimiento son codificadas a través de un arreglo lineal de células referido como esquema perceptual de affordances (APS), el cual representa las posibles rotaciones que la rata puede ejecutar de -180° a $+180^\circ$ en intervalos de 45° a partir de su actual localización y dirección de su cabeza. Cada affordance es representada como una distribución Gaussiana dentro de APS, donde el nivel de actividad de una neurona i se calcula como lo describe la ecuación (3):

$$APS_i = h e^{-\frac{(i-a)^2}{2d^2}}, \quad (3)$$

donde d es el ancho (varianza) de la distribución Gaussiana, h es su altura, y a es su posición media que depende de la rotación particular. Específicamente, el modelo emplea $a = 4 + 9m$ con m un valor entero entre 0 y 8 que corresponde a una rotación entre -180° y $+180^\circ$ en intervalos de 45° . Existe una distribución de Gauss en APS por cada rotación o affordance disponible en un instante de tiempo dado. Por ejemplo, una rata orientada al norte y localizada en el punto de bifurcación de un laberinto tipo T como el mostrado en la Figura III.3(a), es capaz de percibir las affordances -90° , $+90^\circ$ y $\pm 180^\circ$, esto es, la rata puede girar 90° a la izquierda, 90° a la derecha, o dar la media vuelta para regresar. El esquema perceptual APS generado por el modelo en este caso tendría la forma ilustrada en la Figura III.3(b). En otra situación, una rata localizada en el centro de un laberinto radial de 8 brazos como el que presenta la Figura III.3(c), percibe ocho affordances (0° , $\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$ y $\pm 180^\circ$), esto es, la rata puede desplazarse hacia delante, puede girar 45° , 90° ó 135° a la derecha o a la izquierda, o puede dar la media vuelta y regresar. La Figura III.3(d) ilustra la representación de estas ocho affordances dentro del arreglo APS.

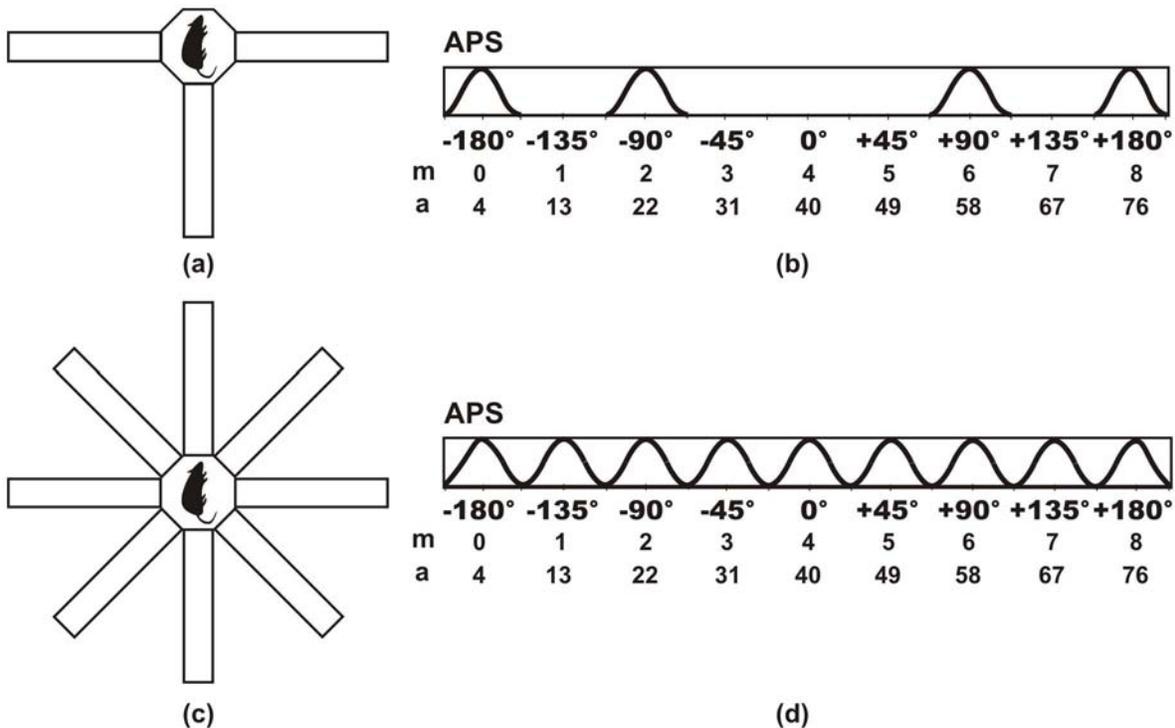


Figura III.3. (a) Diagrama de un laberinto tipo T con la rata orientada al norte en el punto de bifurcación. (b) El esquema perceptual de affordances APS que genera el módulo de procesamiento de affordances del modelo en la situación descrita en (a). (c) Diagrama de un laberinto radial de 8 brazos con la rata ubicada en el punto central. (d) El esquema perceptual de affordances APS generado en el caso descrito en (c). En (b) y (d) se indica la posición media a de cada distribución de Gauss en APS junto con el valor entero m utilizado para calcular a en la ecuación (3).

En el trabajo realizado por Guazzelli y colaboradores (1998), las affordances también son codificadas a través de distribuciones de Gauss dentro de un esquema perceptual, y adicionalmente, este esquema es procesado para producir estados de affordances que representan distintas situaciones motrices experimentadas por el animal dentro de un ambiente y experimento dados. Posteriormente, affordances específicas son reforzadas en dichos estados contribuyendo así a determinar las decisiones de la rata simulada cuando experimenta situaciones motrices particulares. Sin embargo, como el modelo de Guazzelli y colaboradores asume que la rata experimentará únicamente situaciones motrices distintas mientras explora el

ambiente, fracasa cuando el animal ejecuta la misma *affordance* reforzada estando en el mismo estado de *affordances* en dos contextos diferentes.

III.4. PROCESAMIENTO DE INFORMACIÓN KINESTÉSICA

La información kinestésica se refiere a las señales internas generadas por el cuerpo de un individuo durante su locomoción. Como se describió en el Capítulo II de la disertación, estas señales son utilizadas por las ratas para llevar a cabo el proceso de integración de trayectoria, por medio del cual, cada vez que se mueven, actualizan la posición de su punto de partida (el ancla ambiental) con respecto a su posición actual.

En cualquier instante de tiempo t , información kinestésica es inyectada al módulo integrador de trayectoria del modelo. En particular, este módulo recibe las magnitudes de la rotación y de la traslación ejecutadas por el animal en el tiempo $t-1$.

El módulo se compone de una capa que integra la trayectoria (DRL), y una capa que genera y detecta patrones de integración de trayectoria (PIFDL). DRL representa la posición del ancla en el ambiente de la rata en un arreglo bidimensional de neuronas referido como esquema perceptual de remapeo dinámico (DR). El uso de la técnica de remapeo dinámico en el proceso de integración de trayectoria fue concebido originalmente por Dominey y Arbib (1992), e implementado por Guazzelli y colaboradores (1998).

La posición del ancla es codificada como una distribución de Gauss en DR, donde el nivel de actividad de una neurona i, j se calcula de acuerdo a la ecuación (4):

$$DR_{i,j} = h e^{-\frac{(i-y)^2}{2d^2} - \frac{(j-x)^2}{2d^2}}, \quad (4)$$

donde d es el ancho de la distribución Gaussiana, h es su altura, y x, y codifican las coordenadas iniciales del ancla en un plano que representa el ambiente particular.

Cada vez que la rata se mueve, la posición del ancla (esto es, la curva Gaussiana) es desplazada en DR en la misma magnitud pero en dirección opuesta al movimiento de la rata, simulando de esta forma cómo el ancla se aleja de la rata conforme la rata se desplaza hacia delante en su ambiente. La Figura III.4(a) ejemplifica, paso a paso, el caso de una rata explorando un laberinto tipo T desde la base de la T (etiquetada como "O") hasta el final del brazo derecho de la T (etiquetado como "E"). La Figura III.4(b) presenta a DR visto desde arriba ilustrando el desplazamiento de la posición del punto de partida de la rata cada vez que ésta se mueve. Cuando la rata inicia la exploración en "O", DR presenta la posición del ancla en sus coordenadas iniciales (2, 4). Mientras la rata está orientada al norte y se mueve paso a paso a lo largo del corredor vertical de la T hasta llegar al punto de bifurcación, la posición del ancla es desplazada en DR renglón por renglón en dirección sur de las coordenadas (2, 4) a (5, 4). Posteriormente, la rata gira a la derecha en el punto de bifurcación de la T orientándose al este, y se mueve paso a paso hasta alcanzar el final del corredor derecho. Entonces, la posición del ancla en DR es desplazada columna por columna en dirección oeste de las coordenadas (5, 4) a (5, 2).

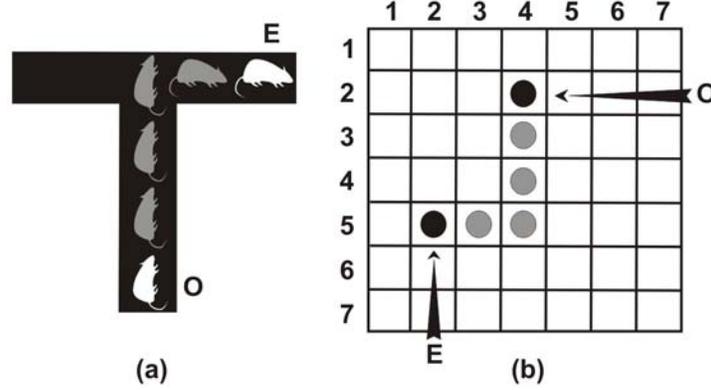


Figura III.4. (a) El caso de una rata explorando un laberinto T paso a paso desde la localidad etiquetada como “O” hasta la localidad indicada como “E”. (b) Vista superior simplificada de un esquema perceptual de remapeo dinámico (DR) representando, cada vez que la rata se mueve, el desplazamiento de la posición del ancla de las coordenadas (2, 4) a (5, 2) en la misma magnitud pero en dirección opuesta al movimiento del animal. Los círculos rellenos representan el pico de la curva de Gauss en DR visto desde arriba.

DRL actualiza la posición del ancla en DR aplicando una operación de convolución entre DR y la “máscara” M , un arreglo bidimensional utilizado para codificar la dirección opuesta de la cabeza de la rata. La operación de convolución es almacenada en un arreglo bidimensional C con las mismas dimensiones de DR. El cálculo de cada elemento k, l de C es descrito por la ecuación (5):

$$C_{k,l} = \sum_{p=-n/2}^{n/2} \sum_{q=-n/2}^{n/2} \left(M_{p+n/2, q+n/2} \times DR_{k-p, l-q} \right), \quad (5)$$

donde n es la dimensión de M . Después, DR es actualizado de acuerdo a C centrando la posición media de la distribución Gaussiana en las coordenadas (r, c) del máximo valor almacenado en C . De esta forma, el nivel de actividad de cada neurona i, j en DR es recalculado según la ecuación (6):

$$DR_{i,j} = h e^{-\frac{(i-r)^2}{2d^2} - \frac{(j-c)^2}{2d^2}} \quad \left| \quad C_{r,c} = \max(C_{k,l}) \quad \forall k,l. \quad (6)$$

En el módulo integrador de trayectoria, la capa DRL está conectada a través del esquema DR a la capa PIFDL. La Figura III.5 ilustra la arquitectura del módulo. En esta versión básica del modelo, cada neurona en DR está conectada aleatoriamente al 50% de las neuronas en PIFDL. No obstante, como una extensión futura del modelo, este alto nivel de proyección será ajustado al nivel conocido o supuesto por los neurocientíficos entre la corteza parietal posterior y la corteza retrosplenial.

Los pesos de conexión entre capas son inicializados aleatoriamente y normalizados entre 0 y 1 de acuerdo con la ecuación (7):

$$w_{ij} = \frac{w_{ij}}{\sum_i w_{ij}}, \quad (7)$$

donde w_{ij} es el peso de conexión entre una neurona i en DR y una neurona j en PIFDL.

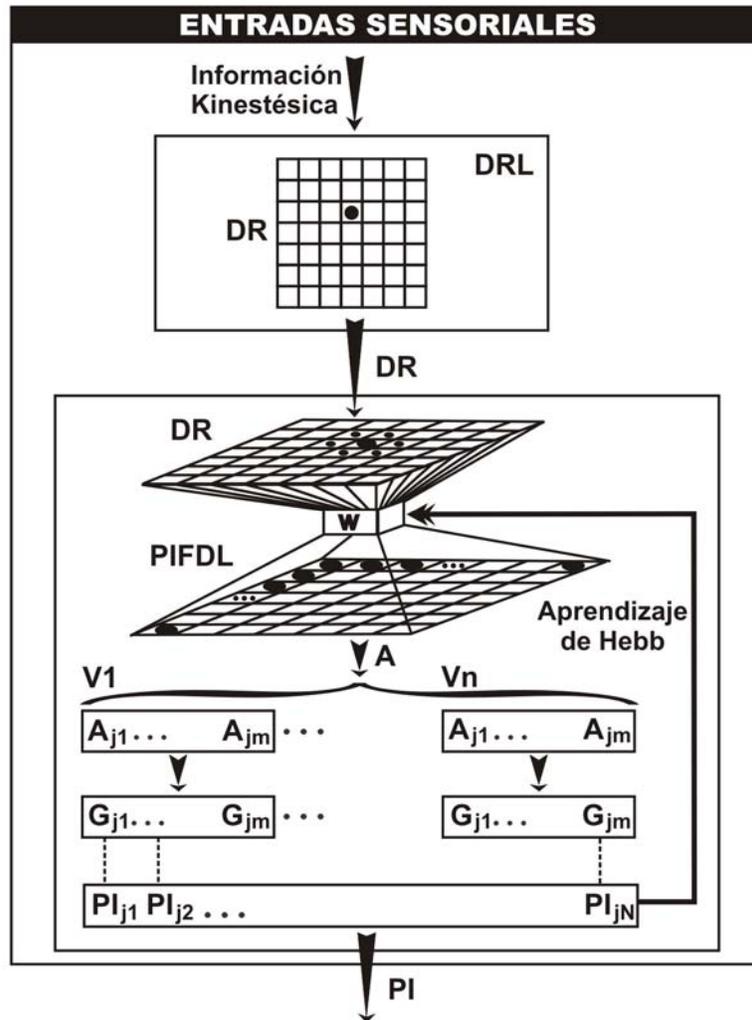


Figura III.5. El módulo integrador de trayectoria del modelo de cognición espacial y navegación en ratas. Glosario: DRL – capa de remapeo dinámico; PIFDL – capa de detección de patrones de integración de trayectoria. Los pesos de conexión w entre el esquema perceptual de remapeo dinámico DR y PIFDL son actualizados a través de la regla de aprendizaje de Hebb. Los valores de activación A de las neuronas j en PIFDL son organizados en vecindarios (V_1, \dots, V_n) de m células cada uno. Nuevos valores de activación G son asignados a todas las neuronas dentro de los arreglos. Tales valores son almacenados en un arreglo lineal PI que representa un patrón de información kinestésica producido por el módulo.

El nivel de actividad A_j de la neurona j en PIFDL se calcula sumando los productos entre cada valor de entrada I_i procedente de la neurona i en DR y el correspondiente peso de conexión w_{ij} como sigue:

$$A_j = \sum_i I_i w_{ij} . \quad (8)$$

La eficacia sináptica entre DRL y PIFDL es actualizada a través de la regla de aprendizaje de Hebb (1949), la cual permite asegurar que la próxima vez que el mismo o un patrón de actividad similar se presente en DRL, el mismo conjunto de unidades en PIFDL se active representando la detección de un patrón específico de información kinestésica. De hecho, la misma regla de Hebb es empleada en los módulos de procesamiento de landmarks y de representación de lugares del modelo, con el fin de posibilitar la generación y detección de patrones de información visual, así como el reconocimiento de lugares.

En el modelo, la aplicación de la regla de Hebb está inspirada en el trabajo original de Guazzelli y colaboradores (1998). Ésta implica la división de la población de N neuronas en PIFDL en un número n de

vecindarios (se utiliza $n=5$) con igual cantidad m de células, donde $m < N$. Las M neuronas más activas dentro de cada vecindario (se utiliza $M=20$) son organizadas lógicamente de acuerdo a su nivel de actividad en el tiempo t , e identificadas por su posición k dentro de la jerarquía de esta forma:

$$\text{ConjuntoOrdenado}(A_{jk}) \mid A_{jk}(t) \geq A_{jk+1}(t), 1 \leq k \leq M .$$

Estas neuronas jk son asociadas con nuevos valores de activación G_{jk} :

$$G_{jk} = (M - k + 1) / M \mid 1 \leq k \leq M , \quad (9)$$

donde G_{jk} toma valores entre 1 y $1/M$ en decrementos de $1/M$. Las restantes $m-M$ neuronas j dentro del vecindario son asociadas con el valor de activación $G_j = 0$.

Los nuevos valores de activación G_j de todas las neuronas j en PIFDL son almacenados en un arreglo lineal de células referido como PI :

$$PI_j = G_j \mid 1 \leq j \leq N . \quad (10)$$

Posteriormente, el patrón de información kinestésica PI producido por el módulo es empleado en la regla de aprendizaje de Hebb como lo describe la ecuación (11):

$$\Delta w_{ij} = \alpha I_i w_{ij} PI_j , \quad (11)$$

donde Δw_{ij} es el cambio en la fuerza de conexión w_{ij} , α es la tasa de aprendizaje, e I_i es el valor de entrada procedente de la neurona i en DR. Como se muestra en (11), la fuerza de conexión entre las neuronas en DR y las neuronas en PIFDL cuyo valor de activación almacenado en PI es 0, no es incrementada, mientras que la fuerza de conexión con neuronas en PIFDL cuyo valor de activación en PI está entre 1 y $1/M$, es incrementada proporcionalmente. Los pesos actualizados son después normalizados entre 0 y 1 utilizando la ecuación (7).

III.5. PROCESAMIENTO DE INFORMACIÓN VISUAL

La información espacial utilizada por las ratas para navegar en el ambiente incluye la localización de la meta con respecto a señales visuales o landmarks. En los ambientes robóticos empleados durante nuestros experimentos, hemos utilizado cilindros como landmarks. Específicamente, en las imágenes percibidas por el robot cuando mueve su cabeza hacia distintas direcciones estando en una localidad dada, los colores de los cilindros son considerados para determinar el tamaño aparente de los landmarks y estimar así la distancia y orientación relativa de cada landmark visible desde la posición del robot. El Capítulo IV profundiza en el cálculo de dichas estimaciones. La Figura III.6 muestra la arquitectura del módulo de procesamiento de landmarks del modelo.

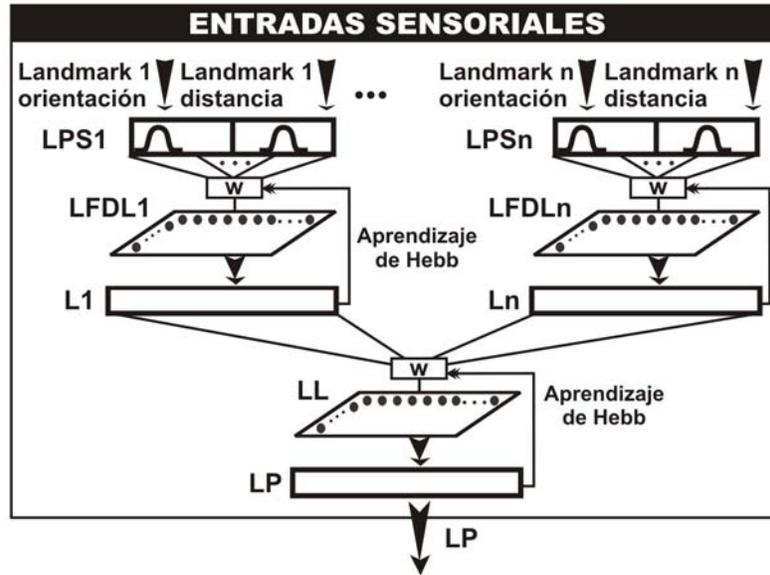


Figura III.6. El módulo de procesamiento de landmarks del modelo de cognición espacial y navegación en ratas. Glosario: LPS – esquema perceptual de landmark; LFDL – capa de detección de patrones de landmark; LL – capa integradora de landmarks. Los pesos de conexión w entre cualquier pareja de capas LPS – LFDL son actualizados a través de la regla de aprendizaje de Hebb. Los patrones de información de landmarks $L1, \dots, Ln$ son integrados por LL para producir un patrón de landmarks global LP que representa una vista egocéntrica desde el animal.

La distancia y orientación relativa de cada landmark hacia la rata son representadas de forma independiente mediante dos arreglos lineales de células incluidos dentro de un esquema perceptual de landmark (LPS). Este tipo de codificación separada de las distancias y orientaciones hacia landmarks también es utilizada por (Guazzelli et. al, 1998).

En nuestro modelo, cada arreglo lineal PS representa la distancia o la orientación en término de una distribución Gaussiana como lo muestra la ecuación (12):

$$PS_i = h e^{-\frac{(i-a)^2}{2d^2}}, \quad (12)$$

donde d es el ancho de la curva de Gauss, h es su altura, y a corresponde a su posición media.

Como en el caso de las distribuciones de Gauss codificadas en el esquema perceptual de affordances (APS), la posición media a de la distribución que representa la orientación del landmark corresponde a la rotación relativa de la rata hacia el landmark, cuyo valor puede encontrarse entre -180° y $+180^\circ$ en intervalos de 45° . Por otro lado, la posición media a de la distribución que representa la distancia hacia el landmark está determinada por la proporción entre la estimación de la distancia actual y la distancia máxima conocida hacia el landmark.

El mismo patrón de conectividad empleado entre las capas DRL y PIFDL del módulo integrador de trayectoria es utilizado entre LPS y una capa particular que detecta patrones de landmark (LFDL). El nivel de actividad A_j de una neurona j en LFDL es calculado a través de la suma de los productos entre cada valor de entrada I_i procedente de la neurona i en LPS y el correspondiente peso de conexión w_{ij} como lo describe la ecuación (8). La eficacia sináptica se mantiene por medio de la regla de aprendizaje de Hebb como se muestra en (11), utilizando en este caso el arreglo lineal de células referido genéricamente como L (en lugar de PI usado en el módulo integrador de trayectoria), y considerando I_i como el valor de entrada procedente de la neurona i en LPS.

En nuestro modelo, como en el de (Guazzelli et al., 1998), una capa LPS conectada a una capa LFDL es agregada por cada landmark presente en el ambiente de esta forma: LPS1–LFDL1, LPS2–LFDL2, ..., LPSn–LFDLn. Posteriormente, todas las capas LFDL son combinadas en una sola capa de landmarks (LL) observando el mismo patrón de conectividad existente entre cualquier pareja LPS – LFDL. El nivel de activación A_j de cada neurona j en LL se calcula de acuerdo con la ecuación (13):

$$A_j = \sum_i L1_i w_{ij} + \sum_p L2_p w_{pj} + \dots + \sum_q Ln_q w_{qj} , \quad (13)$$

donde $L1_i$, $L2_p$, Ln_q son los valores de activación procedentes de las neuronas i , p , q en LFDL1, LFDL2, LFDLn, y w_{ij} , w_{pj} , w_{qj} son los correspondientes pesos de conexión entre las neuronas i , p , q en LFDL1, LFDL2, LFDLn y la neurona j en LL.

La regla de aprendizaje de Hebb mostrada en (11) actualiza los pesos de conexión entre las capas LFDL1, LFDL2, LFDLn y LL, produciendo grupos de neuronas en LL que responden ante patrones de información específicos derivados de la integración de todos los landmarks presentes en el ambiente de la rata. De esta forma, cualquier patrón de información visual almacenado en el arreglo referido como LP representa una vista egocéntrica desde la ubicación y orientación actuales del animal.

III.6. REPRESENTACIÓN Y RECONOCIMIENTO DE LUGARES

La estructura del módulo de representación de lugares del modelo se muestra en la Figura III.7.

Como se describió en la Sección II.3, muchos modelos combinan información kinestésica y visual para determinar la actividad de las células de lugar del hipocampo. El modelo propuesto también emplea dicha integración al nivel de la capa de células de lugar PCL. Por lo tanto, cada neurona en la capa PIFDL del módulo integrador de trayectoria del modelo es conectada aleatoriamente al 50% de las neuronas en la capa PCL, al igual que cada neurona en la capa LL del módulo de procesamiento de landmarks. Los pesos de conexión entre capas son inicializados aleatoriamente y normalizados entre 0 y 1. El nivel de activación A_j de cada unidad j en PCL es calculado como se describe en la ecuación (14):

$$A_j = \sum_i PI_i w_{ij} + \sum_q LP_q w_{qj} , \quad (14)$$

donde PI_i es el valor de activación procedente de la neurona i en el patrón de información kinestésica PI producido por PIFDL, LP_q es el valor de activación procedente de la neurona q en el patrón de información de landmarks LP generado por LL, w_{ij} es el peso de conexión entre la entrada i de PIFDL y la unidad j en PCL, y w_{qj} es el peso de conexión entre la entrada q de LL y la unidad j en PCL.

La eficacia sináptica entre capas es mantenida a través de la regla de Hebb mostrada en la ecuación (11), utilizando en este caso los valores de activación almacenados en el arreglo lineal de células referido como PC , y considerando l_i como el valor de entrada procedente de la neurona i en PI o en LP . De esta forma, PC almacena la actividad conjunta registrada por la colección completa de neuronas en PCL, el cual codifica la información kinestésica y la información visual egocéntrica percibidas por la rata estando en la localidad actual orientada hacia una dirección particular. En este sentido, la respuesta de las neuronas en PCL es modelada de manera unidireccional.

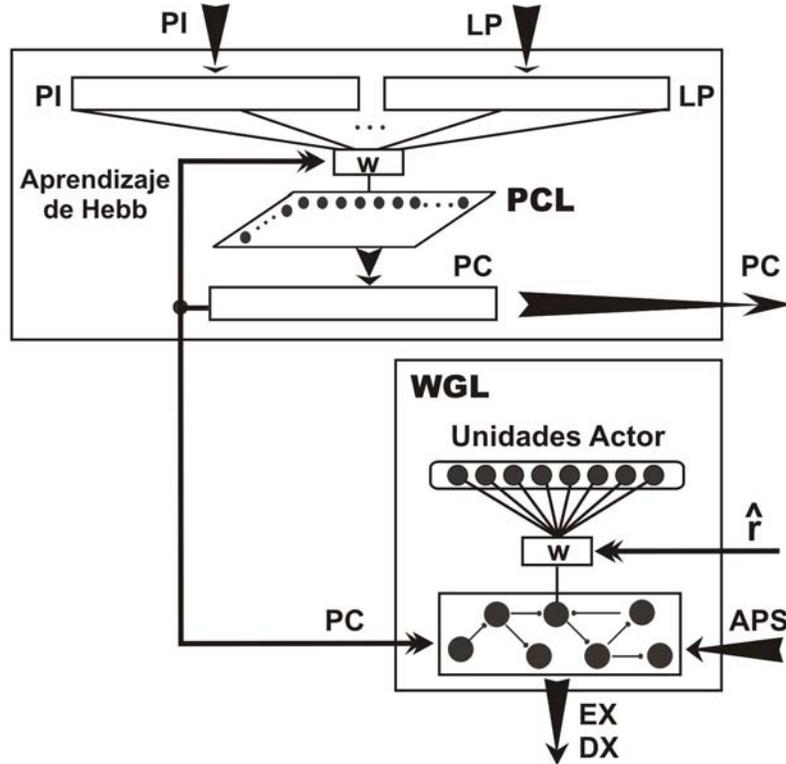


Figura III.7. El módulo de representación de lugares del modelo de cognición espacial y navegación en ratas. Glosario: PCL – capa de células de lugar; WGL – capa del mapa cognitivo; PI – patrón de información kinestésica; LP – patrón de información de landmarks; w – pesos de conexión entre capas; PC – patrón de información de un lugar; APS – esquema perceptual de affordances; \hat{r} – señal de refuerzo efectivo; EX – expectativas de máxima recompensa sobre una secuencia de nodos en el mapa y sus correspondientes direcciones (DX).

El mapa cognitivo es implementado por la capa WGL del modelo. Como en la gráfica del mundo propuesta por (Guazzelli et al., 1998), los nodos del mapa representan lugares distinguibles, y los arcos entre nodos están asociados con la dirección de la cabeza de la rata cuando se mueve de un nodo al siguiente, y con el número de pasos que ejecutó para lograrlo.

El modelo asume que el animal puede orientarse en ocho direcciones estando en cualquier localidad dada, i.e. de 0° a 315° en intervalos de 45° , de acuerdo a un marco de referencia aloentróico que es relativo al punto de partida de la rata en el proceso de exploración. Los ocho patrones de activación generados por la capa PCL son almacenados en unidades Actor. Por lo tanto, cada nodo en el mapa (un lugar) puede ser conectado a ocho unidades Actor (ocho vistas), una por cada dirección.

Para determinar si la rata reconoce o no un lugar en el ambiente, WGL busca el patrón de activación actual PC producido por PCL dentro de las unidades Actor que pertenecen a todos los nodos en el mapa. Esta búsqueda implica el cálculo del grado de similitud SD entre PC y cada patrón almacenado pat :

$$SD = \frac{\sum_{i=1}^N \min(pat_i, PC_i)}{\sum_{i=1}^N PC_i}, \quad (15)$$

donde i es el índice para las neuronas, N es la cantidad total de células en PC o en cualquier pat , y min es una función que determina el valor mínimo en sus dos argumentos.

El modelo propuesto distingue entre dos casos: (a) si existe al menos un SD que excede cierto valor de umbral cercano a 1, la unidad Actor que almacena el patrón de activación con el SD más grande se considera el ganador; (b) si no existe un ganador, WGL crea una nueva unidad Actor asociada a la dirección actual de

la rata para almacenar el patrón *PC*. Posteriormente, WGL activa o crea un nodo en el mapa dependiendo de las siguientes consideraciones:

- Si las affordances percibidas por el animal en el tiempo t son diferentes a las percibidas en el tiempo $t-1$ y una nueva unidad Actor fue creada, WGL crea un nuevo nodo, lo conecta a la nueva unidad Actor, y lo asume como el nuevo nodo activo en el mapa. Este proceso queda ilustrado en la Figura III.8(a).
- Si las affordances no cambiaron de t a $t-1$ y una nueva unidad Actor fue creada, WGL promedia el patrón de activación almacenado en la nueva unidad Actor y el patrón almacenado en la unidad Actor del nodo activo en el mapa que está asociado a la orientación actual de la rata (ver Figura III.8(b)).
- Si existe una unidad Actor ganadora, un arco del nodo activo en el mapa al nodo conectado a dicha unidad Actor se crea en caso de ser necesario, y este nodo se convierte en el nuevo activo (ver Figuras III.8(c-d)).

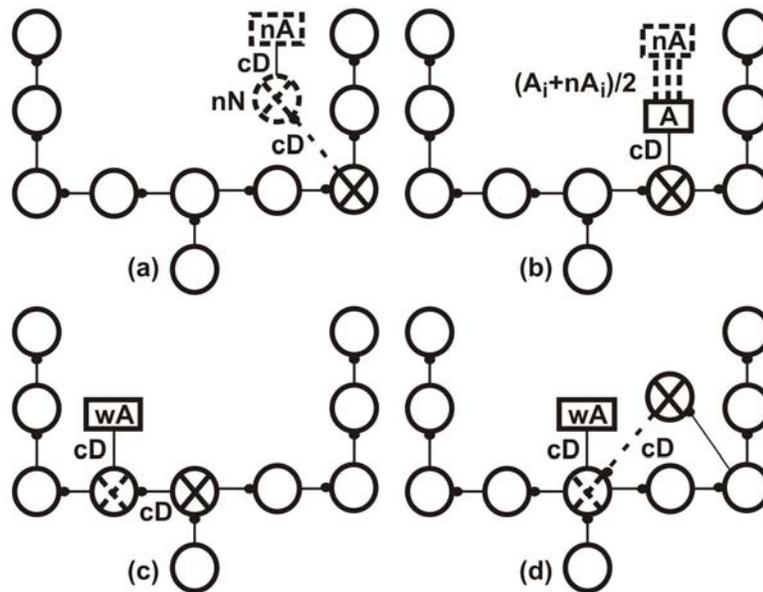


Figura III.8. El proceso de activación/creación de nodos en el mapa cognitivo. Se utilizan líneas punteadas para ilustrar nuevos componentes, y el nodo cruzado representa el nodo activo. Glosario: nA – nueva unidad Actor; nN – nuevo nodo; cD – dirección actual de la rata; A – unidad Actor existente; wA – unidad Actor ganadora. (a) La creación de un nuevo nodo. (b) El promedio entre los patrones de activación de dos unidades Actor. (c) La activación de un nodo existente. (d) La conexión entre dos nodos existentes.

Las unidades Actor conectadas al nodo activo del mapa interactúan de forma competitiva, lo cual contribuye a seleccionar la próxima dirección de movimiento del animal. Para lograrlo, cada conexión es asociada con un peso y un rastro de elegibilidad. El peso representa la expectativa de la rata de obtener una recompensa al orientarse hacia la dirección particular que corresponde a la unidad Actor estando en la localidad actual. El rastro de elegibilidad, por otro lado, marca la conexión como elegible para ser reforzada posteriormente en el tiempo. Al respecto, la Sección III.7 presenta la explicación detallada.

Como en (Guazzelli et al., 1998), nuestro modelo propone la idea de anticipar la próxima dirección hacia la cual el animal debe orientarse para obtener la máxima recompensa estando en la localidad actual. Dicha anticipación implica revisar una secuencia de nodos del mapa (se utilizan tres) empezando por el nodo activo. La Figura III.9 ilustra este análisis considerando el caso del mapa de un laberinto tipo T donde la rata está localizada en la base de la T orientada a 90° (norte), y la recompensa (comida) está colocada al final del corredor izquierdo de la T. Específicamente, WGL analiza los pesos de las unidades Actor conectadas al nodo activo na , y la dirección DX_{na} correspondiente al peso más alto EX_{na} es seleccionada. Después se considera el nodo $na+1$ apuntado por el nodo activo en la dirección elegida DX_{na} , y la dirección DX_{na+1} de

su unidad Actor asociada con el peso más alto EX_{na+1} es seleccionada siempre y cuando este peso sea mayor que EX_{na} . El mismo procedimiento se lleva a cabo para el siguiente nodo $na+2$ en secuencia. Por tanto, los valores de salida de la capa WGL compuestos de las diferentes direcciones DX (tres como máximo) seleccionadas sobre la secuencia de nodos, y de los correspondientes pesos EX , son determinados como lo describe la ecuación (16):

$$\begin{aligned}
 EX_{na}^{da} &= w_{na}^{da}; DX_{na} = da \quad | \quad w_{na}^{da} = \max(w_{na}^{d1}, \dots, w_{na}^{d8}) \\
 EX_{na+1}^{db} &= w_{na+1}^{db}; DX_{na+1} = db \\
 &| \quad w_{na+1}^{db} = \max(w_{na+1}^{d1}, \dots, w_{na+1}^{d8}) \wedge w_{na+1}^{db} > w_{na}^{da} \quad , \\
 EX_{na+2}^{dc} &= w_{na+2}^{dc}; DX_{na+2} = dc \\
 &| \quad w_{na+2}^{dc} = \max(w_{na+2}^{d1}, \dots, w_{na+2}^{d8}) \wedge w_{na+2}^{dc} > w_{na+1}^{db}
 \end{aligned} \tag{16}$$

donde EX_{na}^{da} , EX_{na+1}^{db} , EX_{na+2}^{dc} son los valores de expectativa de recompensa seleccionados a partir de los nodos del mapa na , $na+1$, $na+2$ en las direcciones da , db , dc ; w_{na}^{da} , w_{na+1}^{db} , w_{na+2}^{dc} son los pesos de conexión entre la unidad Actor da y el nodo na , la unidad Actor db y el nodo $na+1$, y la unidad Actor dc y el nodo $na+2$; $d1, \dots, d8$ son las ocho direcciones de 0° a 315° en intervalos de 45° ; y DX_{na} , DX_{na+1} , DX_{na+2} son las direcciones seleccionadas.

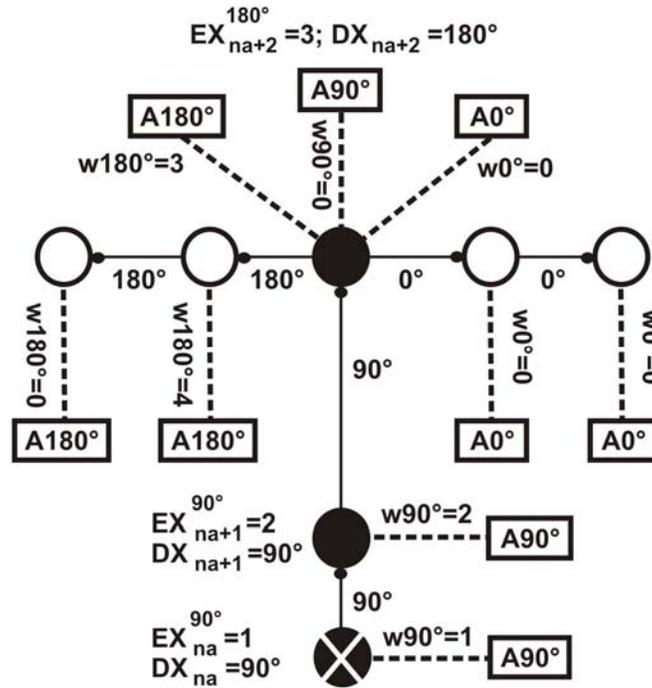


Figura III.9. Ejemplificación del proceso de anticipación de la próxima dirección de la rata considerando el mapa de un laberinto tipo T. El nodo cruzado representa el nodo activo (localidad actual de la rata). En este caso, la rata se encuentra localizada en la base de la T y orientada a 90° (norte), y la recompensa (comida) está colocada al fondo del brazo izquierdo de la T. Los nodos rellenos son los involucrados en el análisis que lleva a cabo la capa WGL del modelo. Se muestran las unidades Actor A conectadas a cada nodo del mapa, así como los pesos de conexión hipotéticos w que les corresponden. Por ejemplo, el nodo activo na está conectado con una fuerza hipotética $w_{90^\circ}=1$ a su unidad Actor A_{90° asociada a la dirección 90° . De acuerdo con los pesos existentes en los nodos na , $na+1$ and $na+2$, WGL seleccionó como resultado del análisis las direcciones 90° y 180° (i.e., $DX_{na} = 90^\circ$, $DX_{na+1} = 90^\circ$ y $DX_{na+2} = 180^\circ$), y los correspondientes valores de expectativa de recompensa 3 y 3 (i.e., $EX_{na}^{90^\circ} + EX_{na+1}^{90^\circ} = 1 + 2 = 3$, $EX_{na+2}^{180^\circ} = 3$).

III.7. APRENDIZAJE

La habilidad de aprender y desaprender localidades de recompensa es atribuida a la rata implementando aprendizaje por refuerzo a través de una arquitectura Actor-Crítico (Barto, 1995), la cual procesa valores esperados de futura recompensa por medio de sus componentes: un Crítico Adaptivo (AC) y un número de unidades Actor, el cual depende de la cantidad de nodos en el mapa. La Figura III.10 ilustra la estructura de los módulos involucrados en el proceso de aprendizaje por refuerzo.

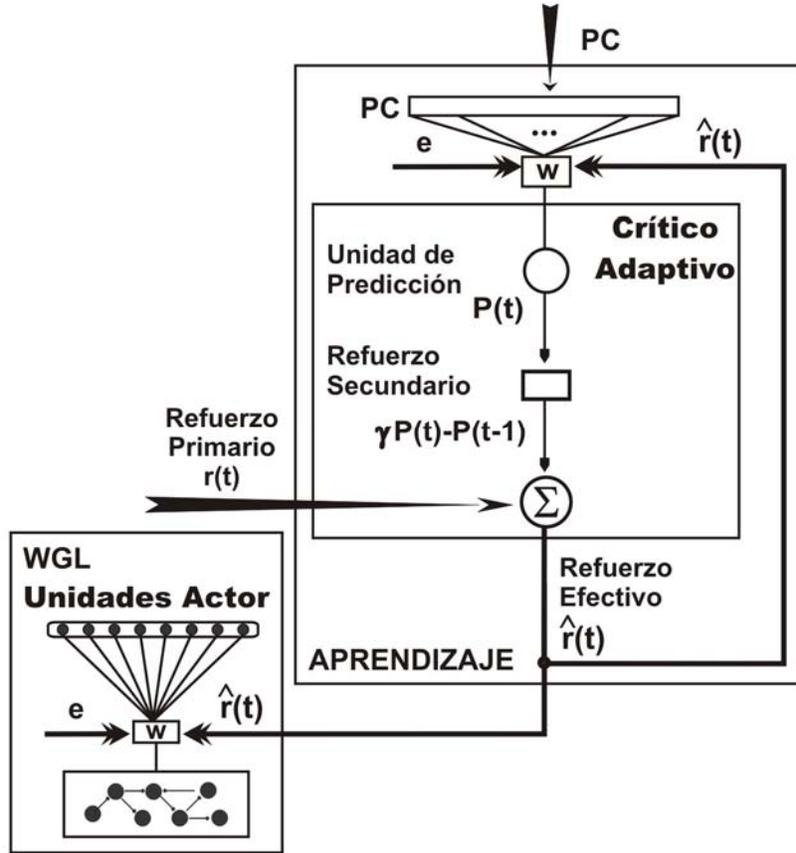


Figura III.10. El módulo de aprendizaje del modelo de cognición espacial y navegación en ratas. Glosario: *PC* – patrón de activación actual en la capa PCL; *w* – pesos de conexión entre capas; *e* – rastros de elegibilidad de las conexiones. $P(t)$ y $P(t-1)$ corresponden a las predicciones del valor futuro del patrón de actividad *PC* en el tiempo t y $t-1$ respectivamente.

AC incluye una unidad de predicción (PU) que estima el valor de recompensa futura de cualquier localidad en un tiempo dado. Para ello, cada neurona en PCL es conectada a PU, y cada conexión i es asociada con un peso w y un rastro de elegibilidad e . En cada paso de tiempo t en un ensayo de un experimento dado, PU calcula el valor futuro P del patrón de actividad *PC* generado por PCL de acuerdo a la ecuación (17):

$$P(t) = \sum_{i=1}^N w_i(t) PC_i(t), \quad (17)$$

donde N es la cantidad total de valores de activación en *PC*.

AC utiliza las predicciones calculadas en los tiempos t y $t-1$ para determinar el refuerzo secundario, descontando la predicción actual a una tasa γ para obtener su valor presente. La suma del refuerzo

secundario con el refuerzo primario o recompensa inmediata r calculada por el módulo de motivación del modelo, constituye el refuerzo efectivo \hat{r} como lo muestra la ecuación (18):

$$\hat{r}(t) = r(t) + \gamma P(t) - P(t-1). \quad (18)$$

El refuerzo efectivo es empleado para actualizar los pesos de conexión w entre la capa PCL y AC, y también entre las unidades Actor y los nodos del mapa. En el primer caso se utiliza la ecuación (19):

$$w_i(t+1) = w_i(t) + \beta \hat{r}(t) e_i(t), \quad (19)$$

donde β es la tasa de aprendizaje. En el segundo caso se usa la ecuación (20):

$$w_k^d(t+1) = w_k^d(t) + \beta \hat{r}(t) e_k^d(t) \quad \forall \text{ node } k, \text{ actor } d, \quad (20)$$

donde $w_k^d(t+1)$ es el peso de conexión entre el nodo k y la unidad Actor que corresponde a la dirección d en el tiempo $t+1$, y $e_k^d(t)$ es el rastro de elegibilidad de la unidad Actor d en el nodo k en el tiempo t .

Como lo muestran las ecuaciones (19) y (20), ambas reglas de aprendizaje dependen de la elegibilidad de las conexiones. Al inicio de cualquier ensayo de un experimento dado, los rastros de elegibilidad en AC y en las unidades Actor son inicializados en 0. En cada tiempo t dentro de un ensayo, los rastros de elegibilidad en AC son incrementados en las conexiones entre PU y la neurona más activa dentro de cada vecindario de la cada PCL. Si la acción ejecutada por la rata en el tiempo $t-1$ le permitió percibir la meta, los rastros de elegibilidad son aumentados más que en el caso opuesto. El rastro de elegibilidad e de cada conexión i en el tiempo t es actualizada entonces según la ecuación (21):

$$e_i(t) = e_i(t-1) + \chi PC_i(t) \quad | \quad PC_i(t) = 1, \quad (21)$$

donde χ es el parámetro de incremento, y PC es el arreglo lineal producido por PCL para almacenar los valores de activación entre 0 y 1 de sus neuronas.

De igual forma, en el tiempo t , el rastro de elegibilidad de la conexión entre el nodo activo na en el mapa y la unidad Actor que corresponde a la dirección actual de la rata dir es incrementado en τ como se describe en (22):

$$e_{na}^{dir}(t) = e_{na}^{dir}(t-1) + \tau. \quad (22)$$

Finalmente, después de actualizar los pesos de conexión entre PCL y AC, y entre las unidades Actor y los nodos del mapa en cualquier instante de tiempo t dentro del ensayo, todos los rastros de elegibilidad decaen a una cierta tasa λ como lo muestra la ecuación (23):

$$e_i(t) = \lambda e_i(t-1). \quad (23)$$

III.8. SELECCIÓN DE ACCIÓN

En cualquier localidad del ambiente, la decisión de la rata referente a girar hacia una dirección específica en un tiempo dado t está determinada en el modelo por el módulo de selección de acción a través de cuatro señales que corresponden a

- las affordances disponibles en el tiempo t (APS),

- una rotación aleatoria entre las affordances disponibles en el tiempo t (RPS),
- rotaciones que no han sido exploradas desde la localidad actual de la rata (CPS), y
- la expectativa global de máxima recompensa (EMR).

Estas señales son representadas en arreglos lineales referidos como esquemas perceptuales por medio de una o más distribuciones de Gauss, cuyas posiciones medias dentro del arreglo corresponden a rotaciones relativas específicas entre -180° y $+180^\circ$ en intervalos de 45° .

El esquema perceptual APS es generado por el módulo de procesamiento de affordances del modelo. En este caso, existen tantas curvas de Gauss dentro del esquema como posibles rotaciones pueda ejecutar la rata en un tiempo dado. Cada curva está centrada en la posición del arreglo que corresponde a la rotación relativa específica.

El esquema perceptual RPS presenta sólo una distribución de Gauss centrada en una posición aleatoria del arreglo elegida entre las posiciones que corresponden a las posibles affordances en el tiempo dado.

El esquema perceptual CPS representa la curiosidad del animal en una localidad dada por ejecutar rotaciones que conducen a lugares aún no representados en su mapa topológico. Por tanto, este esquema presenta tantas curvas de Gauss como rotaciones no hayan sido ejecutadas por el animal estando en la localidad actual dentro de las affordances ahí disponibles en el tiempo dado. Cada Gaussiana es centrada en la posición del arreglo que corresponde a la rotación no ejecutada.

Para construir EMR, el módulo de selección de acción utiliza los valores de expectativa de recompensa EX y sus correspondientes direcciones DX seleccionadas por la capa WGL sobre la secuencia de nodos del mapa a partir del nodo activo. Como se describió en la Sección III.6, pueden existir a lo más tres diferentes direcciones y valores de expectativas de recompensa asociados a los tres nodos en la secuencia. Cada valor de expectativa de recompensa EX_k es representado en EMR como una distribución de Gauss, por lo que el esquema EMR almacena tres curvas como máximo. La posición media de cada curva en el arreglo corresponde a la rotación que debe ejecutar la rata para orientarse a la dirección DX_k , y su altura depende de EX_k . De esta forma, EMR es inicializado como sigue:

$$EMR_i = \sum_k \left(\frac{EX_k}{\max(EX)} e^{-\frac{(i-M_k)^2}{2d^2}} \right), \quad (24)$$

donde d es el ancho de la distribución de Gauss, \max es una función que determina el valor de expectativa de recompensa más grande entre todos los valores EX , y M_k corresponde a la posición media de la curva Gaussiana para EX_k . M_k se define como $4+9m$ con m un valor entero entre 0 y 8, el cual corresponde a la rotación específica entre -180° y $+180^\circ$ en intervalos de 45° que la rata debe ejecutar para orientarse a la dirección DX_k . La Figura III.11(a) ilustra la inicialización de EMR considerando el caso del mapa presentado en la Figura III.9 y los valores EX y DX seleccionados por WGL.

Para generar la señal de expectativa global de recompensa que influirá en el próximo comportamiento de la rata, el "centro de masa" c es calculado sobre el esquema EMR, considerando únicamente los valores de activación de las neuronas en EMR localizadas en la posición media de las diferentes distribuciones Gaussianas. El cálculo del centro de masa queda descrito en la ecuación (25):

$$c = \text{int} \left(\frac{\sum_i \left(EMR_i \times \left(i - \frac{nPS}{2} \right) \right)}{\sum_i EMR_i} \right) + \frac{nPS}{2}, \quad (25)$$

donde c corresponde a la posición del centro de masa en EMR, EMR_i es el valor de activación de la neurona localizada dentro del arreglo en la posición media i de una de las curvas de Gauss existentes en EMR, nPS es la cantidad total de neuronas en el esquema EMR, e int es una función que determina el valor entero de su argumento. Posteriormente, el esquema EMR es actualizado para representar la señal de expectativa global de recompensa por medio de una sola curva de Gauss centrada en c con altura correspondiente a la suma de las alturas de cada curva k preexistente en EMR:

$$EMR_i = \left(\sum_k \frac{EX_k}{\max(EX)} \right) e^{-\frac{(i-c)^2}{2d^2}}. \quad (26)$$

En caso de que no exista una affordance disponible que coincida con el centro de masa calculado sobre EMR, este centro es desplazado en el arreglo a la posición que corresponde a la rotación relativa que orienta a la rata hacia la dirección DX_{na} seleccionada a partir del nodo activo na (el primer nodo en la secuencia analizada por WGL). La Figura III.11(b, c) ilustra la determinación del centro de masa para el ejemplo provisto por la Figura III.9, y la actualización aplicada a EMR.

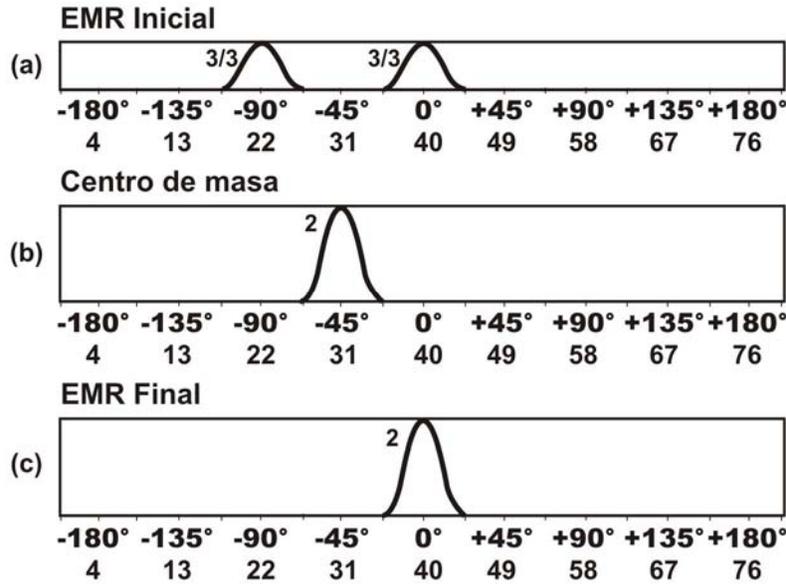


Figura III.11. Ejemplificación del proceso llevado a cabo por el módulo de selección de acción del modelo para generar el esquema perceptual EMR que representa la expectativa global de máxima recompensa. (a) Considerando los valores de expectativa de recompensa $EX_{na}^{90^\circ} = 1$, $EX_{na+1}^{90^\circ} = 2$, $EX_{na+2}^{180^\circ} = 3$ en las direcciones $DX_{na} = 90^\circ$, $DX_{na+1} = 90^\circ$, $DX_{na+2} = 180^\circ$ seleccionadas por WGL en el mapa presentado en la Figura III.9 cuando la rata está orientada a 90° y localizada en la base de un laberinto tipo T, EMR es inicializado con dos curvas de Gauss, una para cada dirección diferente DX . La distribución Gaussiana de altura 1 asociada con la dirección 90° está centrada en la posición 40 del arreglo EMR, la cual corresponde a la rotación relativa 0° , mientras que la curva Gaussiana 180° también de altura 1 está centrada en la posición 22 correspondiente a la rotación relativa -90° . (b) EMR es actualizado para presentar sólo una distribución de Gauss de altura 2 y posición media 31, la cual resulta del cálculo del centro de masa utilizando la ecuación (25). (c) Aún cuando la rotación relativa -45° (posición 31 en EMR) anticipa la dirección objetivo 135° desde la localidad actual de la rata (dado que la recompensa está colocada al fondo del brazo izquierdo de la T), ésta no coincide con una affordance posible desde ahí, así que la posición media de la Gaussiana es desplazada a la posición 40 que corresponde a la rotación 0° y a la dirección $DX_{na} = 90^\circ$.

Finalmente, el módulo de selección de acción suma, neurona por neurona, los valores de activación almacenados en los esquemas perceptuales APS, RPS, CPS y EMR produciendo un nuevo esquema perceptual referido como S, donde el valor de activación de cualquier neurona i es calculado a partir de la ecuación (27):

$$S_i = APS_i + RPS_i + CPS_i + EMR_i . \quad (27)$$

La influencia de cada señal en S depende de la altura de sus correspondientes distribuciones Gaussianas. Específicamente, el orden de importancia de tales señales en la selección de la próxima acción a ejecutar por la rata es el siguiente: (i) EMR, (ii) APS, (iii) CPS, y (iv) RPS. La Figura III.12 ilustra la generación de S.

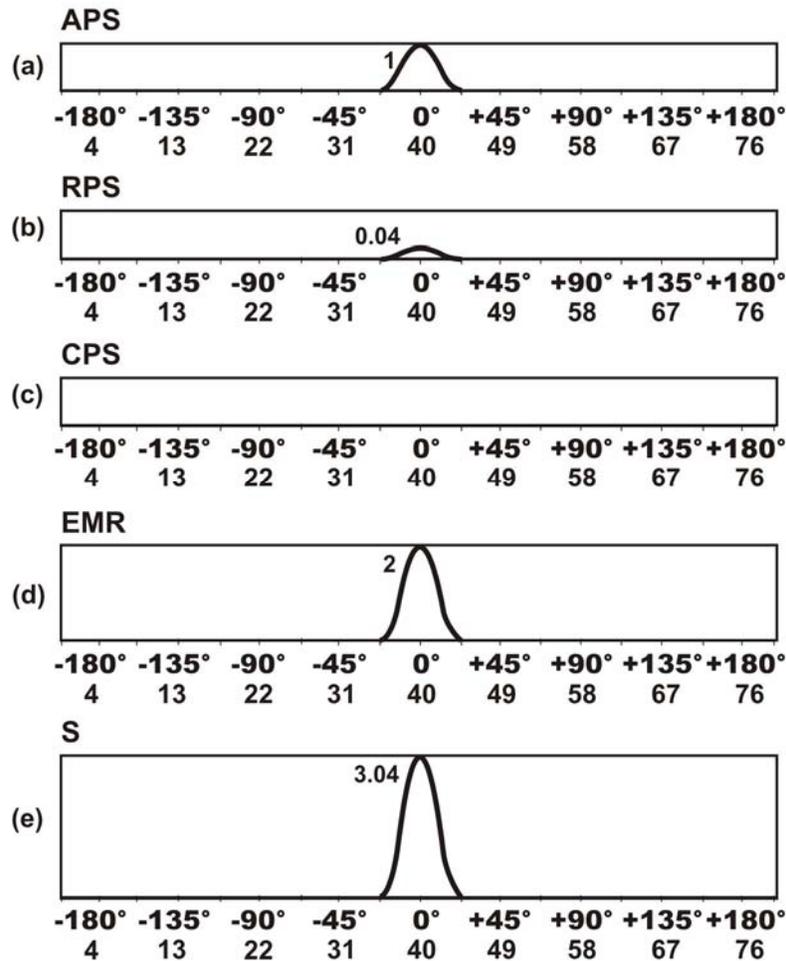


Figura III.12. Ejemplificación del proceso para generar el esquema perceptual S utilizado por el módulo de selección del modelo para determinar la próxima dirección de la rata *DIR*, la rotación requerida *ROT* y el próximo desplazamiento *DIS*, considerando la situación presentada en la Figura III.9. (a) El esquema perceptual de affordances (APS) producido por el modelo cuando la rata está orientada a 90° y localizada en la base de un laberinto tipo T (el animal sólo puede avanzar hacia delante). (b) El esquema perceptual aleatorio (RPS) mostrando una distribución Gaussiana centrada en la posición que corresponde a la única affordance posible. (c) El esquema perceptual de curiosidad (CPS) vacío, pues no existen posibles rotaciones aún no ejecutadas por la rata a partir de su actual localidad. (d) El esquema perceptual de expectativa global de máxima recompensa (EMR) creado en la Figura III.11. (e) El esquema perceptual S resultante de sumar los esquemas perceptuales mostrados en (a), (b), (c) y (d). De acuerdo con S, *ROT* es 0° (i.e., el animal no requiere rotar para orientarse a la próxima dirección), *DIR* es 90°, y *DIS* es “un paso” dado que *DIR* es igual a la dirección actual de la rata.

En el arreglo resultante S, el módulo de selección de acción considera la posición de la neurona que tenga el valor de activación más alto con el fin de determinar la próxima dirección de la rata *DIR*, de 0° a 315° en

intervalos de 45° , y la rotación *ROT* requerida para apuntar hacia tal dirección. Si *DIR* es diferente de la dirección actual, el próximo desplazamiento de la rata *DIS* es 0, dando al animal la oportunidad de obtener una perspectiva diferente del mismo lugar. De otro modo, *DIS* es 1, valor que corresponde a “un paso” en dirección *DIR*.

En el trabajo de Guazzelli y colaboradores (Guazzelli et al., 1998), el proceso de selección de acción considera la adición de las expectativas de recompensa derivadas de su modelo TAM con las expectativas de recompensa producidas por su modelo WG. Sin embargo, como las expectativas de recompensa procedentes de TAM son calculadas asumiendo que el animal sólo experimenta estados de affordances o situaciones motrices diferentes dentro del ambiente, la rata simulada fracasa al tratar de encontrar la meta cuando el laberinto en el que navega incluye dos o más localidades de decisión que ofrecen el mismo estado de affordances.

En nuestro modelo, después de terminar cualquier ensayo de entrenamiento en un experimento dado, la rata regresa a su punto de partida, y el módulo de selección de acción calcula la próxima dirección de la rata *DIR* a partir del mapa construido. En virtud de que el animal marca como visitados los arcos que conectan la trayectoria de nodos seguida durante el ensayo, el módulo de selección de acción emplea las direcciones opuestas de tales arcos para calcular *DIR* durante el proceso de retorno. Mientras la rata regresa al punto de partida, elimina dichas marcas de los arcos.

Para hacer posible la navegación orientada a metas, a diferencia de lo que ocurre en el trabajo de Guazzelli y colaboradores (Guazzelli et al., 1998), el módulo de selección de acción del modelo implementa un proceso de refuerzo hacia atrás sobre los nodos incluidos en la trayectoria seguida por la rata. Así, los rastros de elegibilidad de las unidades Actor son actualizados en la dirección de los arcos que conectan los nodos de la ruta. Cada rastro de elegibilidad es actualizado en una cantidad dada de refuerzo dividida entre la cantidad de pasos que ejecutó la rata para moverse de un nodo al siguiente en la trayectoria. Si el animal encontró la meta al final de la ruta, la actualización es positiva; de otro modo, es negativa. El refuerzo es inicializado en una cantidad específica al comienzo de cualquier ensayo del experimento, y esta cantidad decrece conforme aumenta la distancia desde un nodo hacia la meta o hacia el final de la ruta.

El decremento en el refuerzo se justifica a través de un factor involucrado en el aprendizaje de laberintos referido por Hull (1932) como gradiente de meta. Según dicho factor, los animales tienden a eliminar errores en las localidades de decisión del laberinto cercanas a la meta de manera más temprana que en aquellas encontradas al inicio del laberinto. Concretamente, Hull postuló que el efecto del refuerzo es el máximo en la meta y que éste disminuye progresivamente conforme el animal se desplaza hacia atrás a través del laberinto.

En el modelo, el proceso de refuerzo hacia atrás es llevado a cabo durante el retorno del animal al punto de partida. Específicamente, el algoritmo ejecutado por el módulo de selección de acción en cada paso de la rata durante el proceso de retorno y el refuerzo hacia atrás es presentado en la Tabla III.I. Es importante hacer notar que el proceso de retorno no es una condición necesaria para mantener los rastros de elegibilidad durante el proceso de aprendizaje. De hecho, el retorno fue implementado con el único propósito de ejecutar los experimentos de manera autónoma. La Figura III.13 ilustra los procesos de retorno y de refuerzo hacia atrás empleados en el modelo, considerando nuevamente el caso de una rata explorando un laberinto tipo T desde la base de la misma hasta el final del corredor izquierdo que ofrece una recompensa.

Los valores de los parámetros más importantes utilizados en las ecuaciones del modelo son presentados en el Apéndice A de la disertación.

Tabla III.I. El algoritmo ejecutado por el módulo de selección de acción del modelo durante el proceso de retorno de la rata al terminar cualquier ensayo de entrenamiento en un experimento dado.

<p>Si el proceso de retorno inicia, entonces</p> <ul style="list-style-type: none"> Recordar el nodo del mapa conectado al activo a través del arco marcado como visitado Eliminar la marca del arco Calcular la próxima dirección de la rata como la opuesta del arco <p>De otro modo, si el proceso de retorno no está iniciando, entonces</p> <ul style="list-style-type: none"> Si las affordances no cambiaron del tiempo $t-1$ a t, entonces <ul style="list-style-type: none"> Utilizar la dirección de la rata en el tiempo t como su próxima dirección en $t+1$ De otro modo <ul style="list-style-type: none"> Actualizar el nodo activo con el nodo que lo apunta a través del arco marcado como visitado Dividir la cantidad de refuerzo (R) por la cantidad de pasos asociada al arco marcado como visitado del nodo activo Si la meta fue alcanzada al terminar el ensayo del experimento, entonces <ul style="list-style-type: none"> Sumar R al rastro de elegibilidad de la unidad Actor que corresponde a la dirección del arco marcado como visitado en el nodo activo De otro modo <ul style="list-style-type: none"> Restar R del rastro de elegibilidad de la unidad Actor que corresponde a la dirección del arco marcado como visitado en el nodo activo Disminuir R Si existe un arco marcado como visitado que apunte al nodo activo en el mapa, entonces <ul style="list-style-type: none"> Recordar el nodo conectado al activo a través de dicho arco Eliminar la marca del arco Calcular la próxima dirección como la opuesta del arco
--

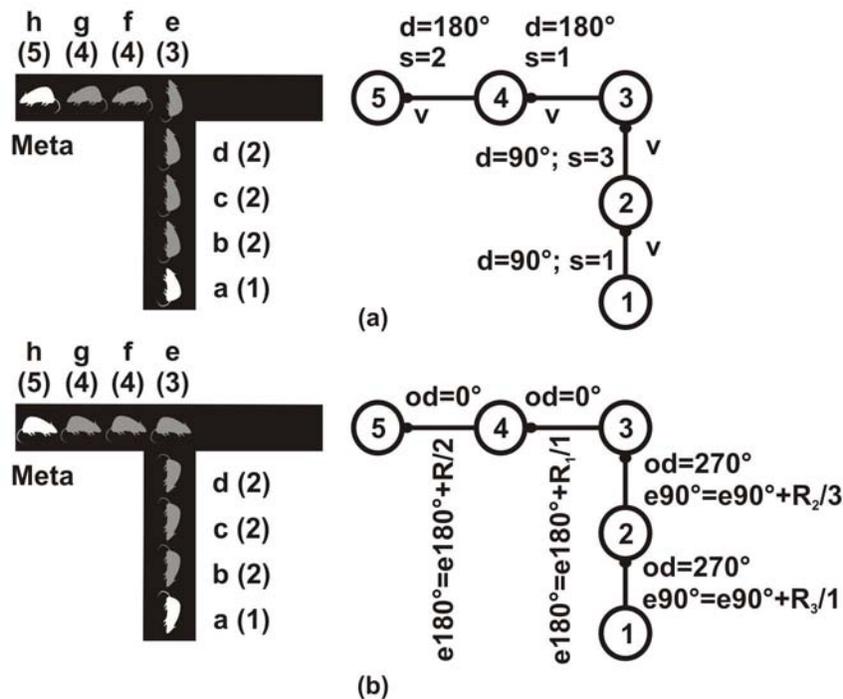


Figura III.13. Ejemplificación de los procesos de retorno y de refuerzo hacia atrás llevados a cabo por el módulo de selección de acción del modelo. (a) La rata explora un laberinto tipo T desde la localidad a hasta la localidad h donde la meta está localizada. En el mapa correspondiente con cinco nodos numerados en orden de creación, los arcos entre nodos están asociados con la dirección de la rata d cuando se desplazó de un nodo al siguiente, y con la cantidad de pasos que ejecutó para lograrlo. Adicionalmente, los arcos en la trayectoria seguida por la rata están marcados como visitados con una "v". (b) La rata regresa de la localidad h a la localidad a eliminando todas las marcas "v" del mapa. En este caso, el mapa muestra las direcciones opuestas od calculadas por el módulo de selección de acción a partir de los arcos en la ruta seguida por el animal. El proceso de refuerzo hacia atrás implica la actualización positiva de los rastros de elegibilidad e en la unidad Actor 180° del nodo 4, la unidad Actor 180° del nodo 3, la unidad Actor 90° del nodo 2, y la unidad Actor 90° del nodo 1. El incremento aplicado está dado por las cantidades de refuerzo decrecientes R , R_1 , R_2 y R_3 divididas por los pasos de la rata 2, 1, 3 y 1 respectivamente.

IV. IMPLEMENTACIÓN DEL MODELO

La implementación del modelo de cognición espacial y navegación en ratas implica tanto el diseño de los módulos que lo componen, como la definición de interfaces que le permitan interactuar con un ambiente virtual o real. El capítulo que se presenta a continuación discute tales aspectos de implementación. En particular, se incluyen tres secciones que describen el uso del Lenguaje de Simulación Neuronal NSL en el diseño de los componentes del modelo, la interacción del modelo con un robot virtual, y la interfaz entre el modelo y el robot real. Estos dos últimos apartados se enfocan a la especificación del procesamiento visual y del control motriz del robot.

IV.1. EL LENGUAJE DE SIMULACIÓN NEURONAL

El modelo de cognición espacial y navegación en ratas fue diseñado e implementado utilizando el Lenguaje de Simulación Neuronal NSL (Weitzenfeld et al., 2002).

NSL ofrece una plataforma basada en tecnología orientada a objetos para construir arquitecturas neuronales (modelado) y ejecutarlas (simulación). El modelado biológico de arquitecturas neuronales tiene como propósito reproducir datos experimentales anatómicos y fisiológicos, lo cual implica la selección de representaciones de datos apropiadas para los componentes neuronales y sus interacciones, la especificación del funcionamiento de los mismos en términos de ecuaciones matemáticas, así como el control de las entradas y parámetros de la arquitectura. Por otro lado, la simulación por computadora de la arquitectura neuronal permite observar su comportamiento ante configuraciones diversas de sus parámetros y entradas generando resultados numéricos y visuales, los cuales son analizados posteriormente para determinar posibles modificaciones al estado actual del modelo.

NSL ha evolucionado desde 1989, cuando el sistema original fue escrito en el lenguaje de programación C. Una segunda versión basada en tecnología orientada a objetos y escrita en C++ fue liberada en 1991. La versión actual NSL 3 incluye modularidad y concurrencia en dos ambientes diferentes, uno en Java y otro en C++. La implementación del modelo propuesto de cognición espacial en ratas se llevó a cabo en la versión NSL 3 para Java.

El sistema de simulación provisto por NSL comprende un subsistema de simulación donde tiene lugar la interacción y el procesamiento de los modelos, y un subsistema de interfaz donde tiene lugar la interacción de gráficas mostradas durante la ejecución de los modelos. Específicamente, las funciones del subsistema de simulación incluyen el ligado del modelo con las librerías provistas por NSL para generar un archivo ejecutable, y la ejecución del modelo y sus módulos en una secuencia específica, mientras que el subsistema de interfaz ofrece librerías gráficas que permiten instanciar marcos, paneles y gráficos, así como controladores de ventana que permiten al usuario interactuar con el proceso de simulación y controlar su ejecución y parámetros.

El modelado biológico de una arquitectura neuronal en NSL implica el diseño de diversos módulos que pueden corresponder a diferentes estructuras del cerebro o representar funcionalidades distintas de una misma estructura. Los módulos en NSL se asocian a objetos que especifican el modelo computacional, y cada módulo puede ser descompuesto en submódulos. En particular, los módulos y submódulos NSL comprendidos dentro del modelo propuesto de cognición espacial y navegación en ratas, su correspondencia biológica, y la interacción entre ellos se presentan en la Tabla IV.I. La funcionalidad interna de cada módulo ha sido descrita en el Capítulo III de la disertación.

Tabla IV.I. Módulos NSL que componen el modelo de cognición espacial y navegación en ratas, y su interacción. Glosario: DR – esquema perceptual de remapeo dinámico; PI – patrón de información kinestésica; LPS – esquema perceptual de un landmark; L – patrón de información de un landmark; LP – patrón de información de landmarks; APS – esquema perceptual de affordances; PC – patrón de información de lugar; EX – expectativas de máxima recompensa y sus direcciones correspondientes (DX).

MÓDULO	SUBMÓDULO	CORRESPONDENCIA BOLÓGICA	ENTRADA – FUENTE	SALIDA – DESTINO
Motivación (MOT)		Hipotálamo lateral	Distancia hacia la meta – Procesamiento visual	Refuerzo primario – ACVTA
Integración de trayectoria	DRL	Corteza parietal posterior	Dirección actual – Control motriz	DR – PIFDL
	PIFDL	Corteza retrosplenial	DR – DRL	PI – PCL
Procesamiento de landmarks	LC	Corteza parietal posterior	Distancias y orientaciones hacia landmarks visibles – Procesamiento visual	LPS – LFDL
	LFDL	Corteza entorhinal	LPS – LC	L – LL
	LL	Corteza entorhinal	L1 – LFDL1 L2 – LFDL2 Ln – LFDLn	LP – PCL
Procesamiento de affordances (AP)		Corteza parietal posterior	Dirección actual – Control motriz Cantidad de pixeles de colores en 3 imágenes – Procesamiento visual	APS – WGL, AS
Representación de lugares	PCL	Hipocampo	PI – PIFDL LP – LL	PC – WGL, ACVS
	WGL	Corteza prefrontal	PC – PCL APS – AP Dirección actual – Control motriz	EX, DX – AS
Aprendizaje	Crítico Adaptivo – VTA (ACVTA)	Área ventral tegmental	Refuerzo primario – MOT Refuerzo secundario – ACVS	Refuerzo efectivo – ACVS, Unidad Actor
	Crítico Adaptivo – VS (ACVS)	Núcleo estriado ventral	PC – PCL Refuerzo efectivo – ACVTA	Refuerzo secundario – ACVTA
	Unidad Actor	Hipocampo	Refuerzo efectivo – ACVTA	
Selección de acción (AS)			APS – AP EX, DX – WGL Dirección actual – Control motriz	Próxima rotación, próxima dirección, desplazamiento – Control motriz

El modelo propuesto es capaz de interactuar con un ambiente virtual o real a través de dos módulos externos de procesamiento visual y de control motriz nombrados en la Tabla IV.I. A nivel simulado, el modelo

utiliza las imágenes provenientes del mundo virtual, y controla la rotación y el desplazamiento del robot virtual dentro de su mundo. Por otro lado, al experimentar con un robot real, el modelo considera las imágenes tomadas por el robot con su cámara local, y lo controla rotándolo y desplazándolo dentro del ambiente físicamente construido.

Las siguientes secciones presentan la interfaz entre el modelo y el robot virtual, y entre el modelo y el robot real, describiendo la funcionalidad específica de los módulos externos de procesamiento visual y de control motriz en cada caso.

IV.2. INTERFAZ MODELO – ROBOT VIRTUAL

El ambiente virtual que se utiliza para simular un experimento particular se describe a través de un archivo XML (*Extensible Markup Language*), en donde se especifica la forma de cada objeto 3D, su color y su posición. El archivo XML es interpretado para dibujar en pantalla los diversos objetos, incluyendo al robot virtual en su localidad de partida.

Se sabe que visualmente las ratas sólo reconocen contrastes y formas. No obstante, la versión actual del sistema propuesto lleva a cabo un procesamiento visual simplificado que se basa únicamente en el reconocimiento de colores. Específicamente, se asignan distintos colores a las paredes del laberinto experimental, al fondo de los pasillos, a la localidad de partida del robot, a la localidad meta, y a landmarks externos, con el fin de determinar las affordances disponibles en cada instante de tiempo, detectar la presencia de la meta, y estimar las distancias y orientaciones hacia landmarks perceptibles desde la posición actual del robot.

En cada paso, el robot virtual toma tres imágenes del ambiente: la primera en la misma dirección actual de su cabeza, la segunda girando la cabeza 90° a la derecha, y la tercera girando la cabeza 90° a la izquierda. Aún cuando el uso de tres imágenes no logra simular el amplio campo visual de 320° que tienen las ratas, ha resultado suficiente durante el procesamiento visual efectuado en los laberintos experimentales hasta ahora utilizados.

El panel que despliega el mundo virtual incluye 240 x 240 píxeles. Cada vez que se lee una imagen del mundo, se toma una muestra de 80 x 80 píxeles y se analiza el color de cada uno. El código numérico que se utiliza para representar dicho color en el contexto del experimento que se esté llevando a cabo, se almacena en una matriz de 80 x 80 códigos que codifica la imagen que el robot virtual percibe en la dirección actual de su cabeza.

El módulo externo de procesamiento visual del sistema considera las tres matrices de códigos y calcula en cada una de ellas la cantidad de píxeles de cada color relevante. Después emplea tales cantidades para realizar las siguientes operaciones:

- Envío de las cantidades de píxeles al módulo de procesamiento de affordances del modelo para determinar las direcciones hacia las que el robot puede avanzar a partir de su localización y orientación actuales. En cada imagen, la cantidad de píxeles del color asociado a las paredes del laberinto y al fondo de los pasillos permite identificar su cercanía.
- Detección de la presencia de la meta. Aún cuando el color asociado a la localidad meta puede estar incluido en las imágenes tomadas por el robot desde diversas distancias, el módulo de procesamiento visual detecta la presencia de la meta únicamente cuando la imagen contiene un número importante de píxeles del color de la meta, el cual corresponde a uno o cero pasos de distancia del robot. Este valor de distancia del robot a la meta es enviado al módulo de motivación del modelo para actualizar el valor del hambre en cada instante de tiempo del experimento dado. En particular, cuando la distancia es un paso

del robot, la percepción de comida se considera como un incentivo, mientras que si es cero pasos, se asume la ingestión de comida por parte del robot, lo cual implica una reducción en el valor del hambre.

- Estimación de las distancias y orientaciones hacia landmarks perceptibles desde la localización y dirección actuales del robot. En los experimentos que se han llevado a cabo, cilindros de diferentes colores dispuestos fuera de los laberintos han sido utilizados como landmarks. El módulo de procesamiento visual cuenta con valores predeterminados del tamaño de los landmarks, los cuales se obtienen, para cada landmark, a partir de la cantidad de píxeles de su color incluidos en la imagen tomada por el robot cuando se encuentra localizado lo más cerca posible del landmark dentro del laberinto. El valor predeterminado del tamaño del landmark es asociado con la distancia entre éste y el robot en términos del número de pasos. Considerando tales valores, la cantidad de píxeles del color de un landmark contenida en la imagen actual tomada por el robot permite determinar el tamaño aparente del landmark y estimar así la distancia hacia la localidad actual del robot. Por otro lado, la orientación del landmark es definida al considerar el ángulo en que el robot haya capturado la imagen que la contiene: 0° , $+90^\circ$ ó -90° . Tanto las distancias como las orientaciones estimadas hacia los landmarks perceptibles son enviadas al módulo de procesamiento de landmarks del modelo, específicamente al submódulo LC presentado en la Tabla IV.I, para codificar dicha información visual a través de esquemas perceptuales de landmarks.

Como lo indica la Tabla IV.I, el módulo de selección de acción del modelo entrega al módulo externo de control motriz los valores referentes a la próxima dirección del robot, a la rotación a ejecutar para orientarse hacia tal dirección, y al desplazamiento a efectuar. En este módulo, el cuerpo del robot virtual y su cabeza son controlados por separado, así que cuando el robot adquiere la nueva dirección indicada por el módulo de selección de acción, lo hace rotando su cuerpo y su cabeza, mientras que cuando captura las tres imágenes del ambiente en cualquier paso de tiempo dentro del experimento dado, lo hace rotando únicamente su cabeza. La magnitud del desplazamiento del cuerpo del robot indicada por el módulo de selección de acción puede ser 0 ó 1. El valor 1 equivale a “un paso” del robot, lo cual ocurre cuando la próxima dirección del robot es igual a su dirección actual. El valor 0 implica que el robot no debe trasladarse, pues su próxima dirección es distinta a la actual y se requiere que perciba el ambiente desde la localidad en que se encuentra, pero desde un ángulo diferente. Las traslaciones y rotaciones del robot virtual en el ambiente son ejecutadas por el módulo de control motriz utilizando la funcionalidad que ofrecen las clases Java3D.

Adicionalmente, el módulo externo de control motriz del sistema propuesto lleva un registro actualizado de la dirección del robot en cada instante de tiempo de un experimento dado, contando para ello con el valor predeterminado de su dirección inicial. Durante el experimento, como lo muestra la Tabla IV.I, la dirección actual del robot es enviada a los módulos de integración de trayectoria, procesamiento de affordances, representación de lugares, y selección de acción, con los propósitos descritos enseguida.

El submódulo DRL del módulo integrador de trayectoria requiere conocer la dirección actual del robot para actualizar la posición de su punto de partida en el esquema perceptual de remapeo dinámico DR. Como se explicó en la Sección III.4 de la disertación, DRL desplaza la posición del ancla en DR en la misma magnitud pero en dirección opuesta al movimiento del robot. De esta forma, cuando la dirección actual indicada por el módulo de control motriz en el tiempo t es diferente de la dirección registrada en $t-1$, las coordenadas del ancla no cambian, mientras que si las direcciones en t y $t-1$ son iguales, las coordenadas son desplazadas una casilla en dirección opuesta a la del robot.

Por otro lado, el módulo de procesamiento de affordances precisa determinar las posibles rotaciones a ejecutar por el robot a partir de su localización y dirección actuales. El submódulo WGL del módulo de representación de lugares necesita asociar los arcos que conectan los nodos del mapa cognitivo con la dirección del robot cuando pasó de un nodo al siguiente. Finalmente, el módulo de selección de acción requiere considerar la dirección actual del robot para calcular la rotación que éste debe efectuar en el próximo instante de tiempo para orientarse hacia su nueva dirección.

IV.3. INTERFAZ MODELO – ROBOT REAL

El modelo corre en línea y en tiempo real utilizando un robot Sony AIBO ERS-210 4-legged y una PC Pentium 4 a 1.8 GHz, la cual se comunica con el robot de manera inalámbrica. Como capacidades sensoriales del robot se emplea únicamente su sistema local de visión 2D, cuyo campo visual cubre 50° en el plano horizontal y 40° en el plano vertical.

Durante los experimentos realizados, tanto la visión como el movimiento del cuerpo y el de la cabeza del robot fueron controlados por medio de una aplicación libre llamada *Tekkotsu* creada y mantenida en la universidad Carnegie Mellon. Se trata de una aplicación basada en tecnología orientada a objetos y en pasaje de eventos, escrita en C++, y diseñada originalmente para ofrecer el manejo de operaciones mecánicas de bajo nivel con el robot Sony AIBO de tal manera que el usuario pueda enfocarse únicamente en la programación de alto nivel. Algunos de los servicios provistos por *Tekkotsu* incluyen procesamiento visual básico, cinemática básica, monitoreo y operación remotos, y soporte para comunicación inalámbrica.

Antes de iniciar cualquier experimento y con el propósito de que el robot reconozca los diferentes colores que podrá encontrar en el ambiente a explorar, el procesamiento visual implica tomar fotografías de tales colores con la cámara local del robot. Dichas fotografías alimentan a un sistema predefinido de calibración de colores, el cual genera como resultado un archivo que incluye la descripción RGB de cada color y un código numérico que le fue asignado. La descripción RGB se refiere a la cantidad de píxeles rojos, verdes y azules que componen al color en cuestión. El archivo con los códigos de colores se almacena en la memoria local del robot.

En cada paso de tiempo t dentro de un experimento dado, el robot Sony AIBO toma tres fotografías no traslapadas de su entorno (0°, +90° y -90°). La lectura de las imágenes captadas por la cámara es llevada a cabo por la aplicación *Tekkotsu*. El resultado de una lectura es una matriz de 176 x 144 píxeles, en donde cada casilla almacena el código numérico asociado con el color del píxel. Al igual que en la situación virtual descrita en la Sección IV.2, el módulo externo de procesamiento visual del sistema propuesto analiza las tres matrices de códigos, en este caso entregadas por *Tekkotsu*, para calcular las cantidades de píxeles de los colores relevantes, y efectuar las operaciones de envío de tales cantidades al módulo de procesamiento de affordances del modelo, detección de la presencia de la meta, y estimación de las distancias y orientaciones hacia landmarks perceptibles desde la localización y dirección actuales del robot.

Como en el caso virtual, el módulo externo de control motriz del modelo recibe del módulo de selección de acción la próxima dirección del robot, la rotación a ejecutar para orientarse hacia tal dirección, y el desplazamiento a efectuar. La aplicación *Tekkotsu* ofrece métodos que implementan las operaciones básicas de traslación y rotación del AIBO, así que el módulo de control motriz se encarga únicamente de utilizar tales métodos, enviando los parámetros que se requieran. Específicamente, el parámetro necesario para ejecutar un desplazamiento es un valor real que define la dirección (positivo – hacia delante, negativo – hacia atrás) y la velocidad (entre -1 y 1) del movimiento. El robot caminará mientras el módulo de control motriz no lo detenga a través del envío de otro comando de desplazamiento con velocidad 0. En el caso de la rotación del cuerpo, el parámetro indica la dirección (positivo – izquierda, negativo – derecha) y el ángulo (entre -1 y 1) del giro. Finalmente, el parámetro de la rotación horizontal de la cabeza puede ser -1 para girar 90° a la derecha, 1 para hacerlo 90° a la izquierda, o 0 para regresarla al frente.

Dado que la entrada kinestésica al modelo no es provista por un odómetro que calcule los desplazamientos lineales y angulares del robot Sony AIBO, las señales de locomoción son registradas por el módulo externo de control motriz, como en el caso del robot virtual. Concretamente, este módulo entrega la dirección actual del robot en cada instante de tiempo de un experimento dado a los módulos de integración de trayectoria, procesamiento de affordances, representación de lugares, y selección de acción.

V. EXPERIMENTACIÓN

Cuatro fueron los aspectos probados durante la experimentación con el robot Sony AIBO: (i) el aprendizaje y desaprendizaje de localidades de recompensa, (ii) la navegación orientada a metas, (iii) la adaptación del mapa cognitivo ante cambios físicos en el ambiente, y (iv) el reconocimiento de lugares en múltiples direcciones. Tales aspectos son discutidos de manera independiente a través de cuatro grandes apartados que dividen este capítulo.

La primera sección introduce los conceptos de cognición espacial relacionados con la discriminación, la discriminación inversa, la extinción de recompensas y la persistencia, como antecedente para presentar una de las tareas espaciales clásicas de discriminación inversa con ratas sanas y ratas con hipocampo lesionado. Adicionalmente, la sección presenta la propuesta de una versión adaptada del modelo de navegación para el caso de ratas lesionadas, y describe los resultados derivados de implementar dicha tarea de discriminación inversa con robots controlados por el modelo de cognición espacial originalmente propuesto, y con robots dirigidos por el modelo de navegación adaptado.

La segunda sección presenta los resultados obtenidos al experimentar con el robot en una tarea de protocolo similar a uno de los experimentos espaciales clásicos de navegación orientada a metas, donde es factible alcanzar la localidad meta desde cualquier otra localidad del laberinto aprendido.

El proceso de adaptación del mapa cognitivo para reflejar cambios en la configuración física de un laberinto, así como los resultados obtenidos de la experimentación con el robot al respecto, quedan descritos en la tercera sección.

Finalmente, la cuarta sección se enfoca al proceso de reconocimiento de lugares que llevan a cabo tanto las ratas como el robot durante la navegación orientada a metas. En particular, se describen los resultados experimentales obtenidos con ratas y con robots resolviendo la misma tarea espacial dentro de un laberinto cíclico, y se presenta el correspondiente análisis conductual comparativo.

V.1. DISCRIMINACIÓN: APRENDIZAJE – DESAPRENDIZAJE DE LOCALIDADES DE RECOMPENSA

El proceso de **discriminación** se refiere a la capacidad del organismo para detectar diferencias (O'Keefe y Nadel, 1978). Los estudios de discriminación en animales tradicionalmente se han enfocado en su habilidad para responder de manera diferencial ante objetos o lugares en el ambiente, pero el énfasis psicológico principal trata sobre el aprendizaje involucrado en la discriminación, esto es, cómo los animales llegan a identificar los aspectos relevantes del ambiente y cómo es que manifiestan comportamientos adaptivos.

La lógica detrás del aprendizaje en una situación de discriminación implica la asociación entre una recompensa y un objeto o lugar. Cuando un animal es introducido por primera vez a tal situación, responde como típicamente lo hace en cualquier ambiente nuevo, con exploración y la construcción de un mapa

cognitivo. Conforme el entrenamiento transcurre, el animal aprende a resolver el problema de discriminación utilizando diferentes hipótesis: la de lugares, referida por O'Keefe y Nadel (1978) como *locale*, y la de orientación o la dirigida por percepción, referidas ambas como *taxon*. Antes de alcanzar un criterio de aprendizaje, los animales sanos emplean la hipótesis de lugares, es decir, aprenden las localidades que ofrecen recompensa. No obstante, después de alcanzar el criterio, si el entrenamiento continúa, los animales responden con base en una hipótesis *taxon*, típicamente de orientación. Este comportamiento implica la repetición de exactamente la misma respuesta por parte del animal aún ante cambios en la recompensa (de tipo, de cantidad o de localidad), situación referida como **resistencia a la extinción de una recompensa o persistencia en la conducta**.

La tarea de discriminación más comúnmente utilizada consiste en la presentación simultánea de dos opciones, recompensando al animal al responder a una de ellas pero no a la otra. Cuando el animal aprende a discriminar entre dos estímulos o lugares eligiendo consistentemente el positivo y evitando el negativo, la situación puede ser alterada de tal forma que la opción previamente positiva se convierte en negativa y viceversa. Esta tarea involucra el problema referido como **discriminación inversa**, el cual implica desaprender una hipótesis previamente correcta y aprender una nueva. Superficialmente, el proceso puede parecer simple dado que los mismos estímulos o lugares se mantienen relevantes durante la tarea inversa, así que un poco de frustración al no recibir la recompensa haría que el animal eligiera la otra opción. Sin embargo, las tareas espaciales de discriminación están relacionadas con el uso de hábitos y un sobreentrenamiento tiene el efecto de prolongar el aprendizaje de la tarea inversa por el cambio de la hipótesis de lugares a la hipótesis de orientación y la consecuente persistencia en la conducta.

Aún considerando lo anterior, los animales sanos son capaces de resolver las tareas espaciales de discriminación inversa. El uso de la hipótesis de orientación que conlleva la persistencia en la conducta se ve alterado cuando la frustración es administrada al animal repetidamente a una tasa rápida, esto es, empujando intervalos de tiempo muy cortos entre ensayos. Dicha alteración provoca que los animales tiendan nuevamente a explorar el ambiente y a depender de la hipótesis de lugares, lo cual implica la adaptación del mapa cognitivo previamente construido en función del nuevo aprendizaje.

En síntesis, la extinción de una recompensa y la persistencia en la conducta del animal son los principales aspectos implicados en el nuevo aprendizaje y en el desaprendizaje o inhibición del viejo aprendizaje. La evaluación apropiada de los efectos de la extinción de una recompensa depende del tipo de hipótesis que el animal haya utilizado durante el aprendizaje. El retiro o cambio de la recompensa es registrado rápidamente en el sistema de mapeo cognitivo llevado a cabo por el hipocampo, lo cual conduce al animal a dejar de elegir la opción que ahora resulta inapropiada. No obstante, en tareas que incorporan hábitos, la persistencia en el comportamiento depende del grado en que la hipótesis de orientación haya sido utilizada. No es típico que el animal cambie su conducta después de un solo ensayo de frustración, pero tampoco ocurre una persistencia absoluta de la conducta habitual, pues el sistema cognitivo del hipocampo interviene para cesar los hábitos al disparar el proceso de exploración. Los animales con hipocampo lesionado abordan los problemas de discriminación de manera distinta, pues están forzados a asociar recompensas con respuestas o procedimientos particulares en virtud de su incapacidad para aprender y desaprender localidades.

Los siguientes apartados presentan la tarea espacial de discriminación inversa implementada por O'Keefe (1983) con ratas sanas y ratas con hipocampo lesionado en un laberinto tipo T y en uno radial de 8 brazos, así como la adaptación del modelo propuesto para el caso de ratas lesionadas, y los resultados obtenidos con robots al experimentar con la tarea inversa y extenderla para un laberinto formado por múltiples Ts.

V.1.1. La Tarea Espacial de Discriminación Inversa de O'Keefe

El proceso de aprendizaje que los animales llevan a cabo al resolver tareas espaciales de discriminación puede involucrar tanto la asociación de recompensas con localidades (hipótesis de lugares), como la de

recompensas con procedimientos (hipótesis de orientación) (O'Keefe y Nadel, 1978). Mientras que la primera hipótesis depende del sistema de mapeo cognitivo provisto por el hipocampo, la segunda depende del sistema de recompensas atribuido al estriado y especifica el comportamiento en términos de rotaciones relativas a la línea media del cuerpo como lo ilustra la Figura V.1(a).

Interesado en demostrar el uso de dichas hipótesis de aprendizaje en ratas sanas y ratas con hipocampo lesionado, O'Keefe (1983) implementó tareas de discriminación y de discriminación inversa en un laberinto tipo T y en uno radial de 8 brazos (Figura V.1(b-c)).

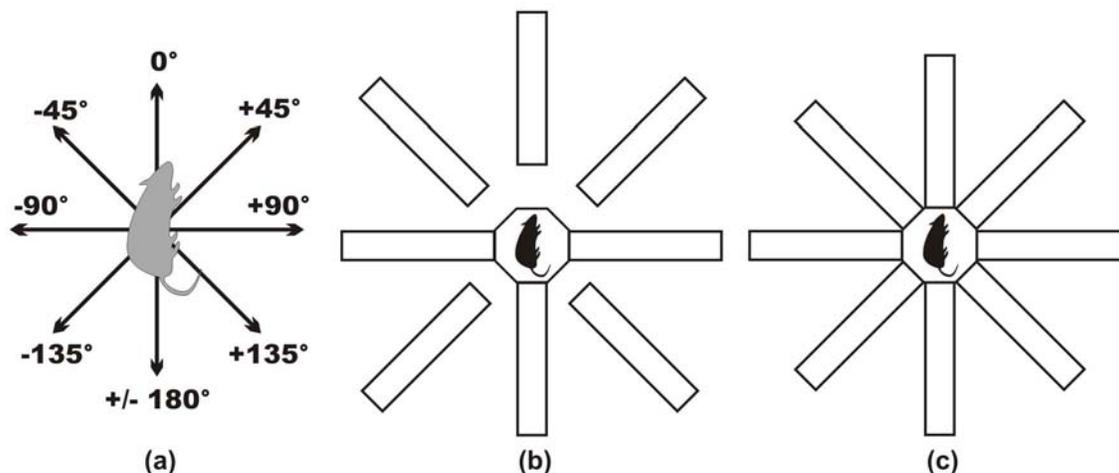


Figura V.1. (a) Sistema de orientación que especifica el comportamiento en términos de rotaciones relativas a la línea media del cuerpo. (b-c) Diagramas de los laberintos empleados por O'Keefe durante las tareas de discriminación y discriminación inversa: un laberinto tipo T resultante de separar cinco brazos de un laberinto radial de 8 brazos (b), y el laberinto radial de 8 brazos (c).

La tarea de discriminación se efectuó en un laberinto radial de 8 brazos sin señales externas que pudieran servir al animal para localizarse a sí mismo. En cada ensayo se retiraron 5 brazos del laberinto de tal manera que los 3 brazos restantes formaban un laberinto tipo T. La recompensa (comida) estaba colocado al fondo de uno de los brazos de la T, y cada 5 ensayos el laberinto era modificado para que otros 3 brazos constituyeran ahora la T. De esta forma, el animal no basaría su elección del brazo correcto en señales externas al laberinto, sino en la hipótesis de lugares o en la hipótesis de orientación. Las ratas con hipocampo lesionado aprendieron la tarea más rápido que las normales, y dos de las seis ratas normales fallaron en alcanzar el criterio de 19 elecciones correctas en 20 ensayos sucesivos dentro de un total de 100 ensayos.

Para determinar el tipo de hipótesis de aprendizaje utilizada por los animales, O'Keefe continuó el experimento después de alcanzar el criterio, intercalando ensayos de prueba en el laberinto de 8 brazos con los ensayos usuales en la T. En caso de que los animales estuvieran basando sus decisiones en la hipótesis de lugares relacionada con la forma de la T dada la ausencia de señales externas, la transición a 8 brazos las haría elegir erróneamente. Sin embargo, los resultados mostraron que tanto las ratas normales como las lesionadas elegían los brazos ubicados en el mismo lado del brazo para el que se les había entrenado en la T. Entonces, lo que las ratas aprendieron fue a rotar a la izquierda o a la derecha. Aún más, algunas ratas no eligieron girar 90°, sino 45° ó 135°. Al respecto, O'Keefe concluyó que la conducta de todas las ratas dependió del uso de la hipótesis de orientación aprendiendo a rotar en la localidad de decisión hacia cualquier ángulo entre 1° y 179°.

Después de entrenar a todos los animales para girar a la izquierda en la T, la tarea de discriminación inversa pretendió evaluar las decisiones de orientación de las ratas al cambiar la localización de la recompensa al fondo del brazo opuesto de la T. Durante esta fase de prueba, O'Keefe intercaló dos intentos en la T con un intento en el laberinto radial de 8 brazos.

Los resultados obtenidos con los animales lesionados indicaron un cambio abrupto en el giro realizado en el punto de decisión de la T del brazo izquierdo al brazo derecho, pero en el laberinto radial, el cambio en la orientación de las ratas era continuo e incremental desde el cuadrante de entrenamiento (-90° , -45°), pasando por seguir derecho (0°) y hasta el cuadrante de prueba ($+45^\circ$, $+90^\circ$). Por otro lado, el desempeño de las ratas sanas en el laberinto tipo T procedió de la misma manera que en las lesionadas, pero su orientación en el laberinto radial no cambió de manera continua e incremental, sino aleatoria.

El resultado de la tarea de discriminación inversa le permitió a O'Keefe ratificar lo siguiente:

- Durante el proceso de aprendizaje, el comportamiento de las ratas con hipocampo lesionado se fundamentó en la hipótesis de orientación, esto es, en la asociación de la recompensa con la respuesta de girar hacia la izquierda en el punto de decisión de la T. La extinción de la recompensa alteró gradualmente la hipótesis de orientación, lo que habilitó a los animales durante el proceso de desaprendizaje para girar a la derecha. No obstante, ambos procesos dependieron de la misma hipótesis.
- Mientras no alcanzaron el criterio de aprendizaje durante el entrenamiento, las ratas sanas utilizaron una hipótesis de lugares. No obstante, después de este evento, basaron el aprendizaje en la hipótesis de orientación al terminar el entrenamiento, y ésta fue alterada abruptamente ante la extinción de la recompensa. Dicha alteración abrupta fue causada por el cambio hacia una hipótesis de localidades durante el proceso de desaprendizaje dada la intervención del hipocampo para disparar nuevamente la exploración del ambiente. La hipótesis de localidades en este caso está relacionada únicamente con la forma del laberinto empleado durante la tarea en virtud de la ausencia de señales externas.

V.1.2. Un Modelo de Navegación para Ratas Lesionadas

Con el fin de reproducir con robots la tarea espacial de discriminación inversa de O'Keefe, el modelo originalmente propuesto de cognición espacial y navegación en ratas fue adaptado para derivar una versión aplicable a animales que presentan lesiones en el hipocampo.

La ausencia de un hipocampo sano en el animal implica la pérdida del sistema de mapeo cognitivo, el cual está determinado por el procesamiento de señales kinestésicas y visuales provistas por la corteza retrosplenial y la corteza entorhinal respectivamente. La adaptación del modelo involucra, entonces, la exclusión de los módulos de integración de trayectoria, procesamiento de landmarks y representación de lugares. Por otro lado, la funcionalidad del módulo de motivación se mantiene, y se modifican sustancialmente los módulos de procesamiento de affordances, aprendizaje y selección de acción. La Figura V.2(a) muestra la arquitectura de esta versión del modelo, y la Figura V.2(b) su posible correspondencia biológica. Las entradas sensoriales incluyen únicamente información relacionada con el estado motivacional interno de la rata y con affordances de movimiento. Las salidas motrices se componen, al igual que en el modelo original, de la próxima dirección de la cabeza de la rata, la rotación requerida para apuntar hacia tal dirección, y el próximo desplazamiento del animal. Las siguientes secciones describen con detalle los módulos del modelo original que han sido modificados.

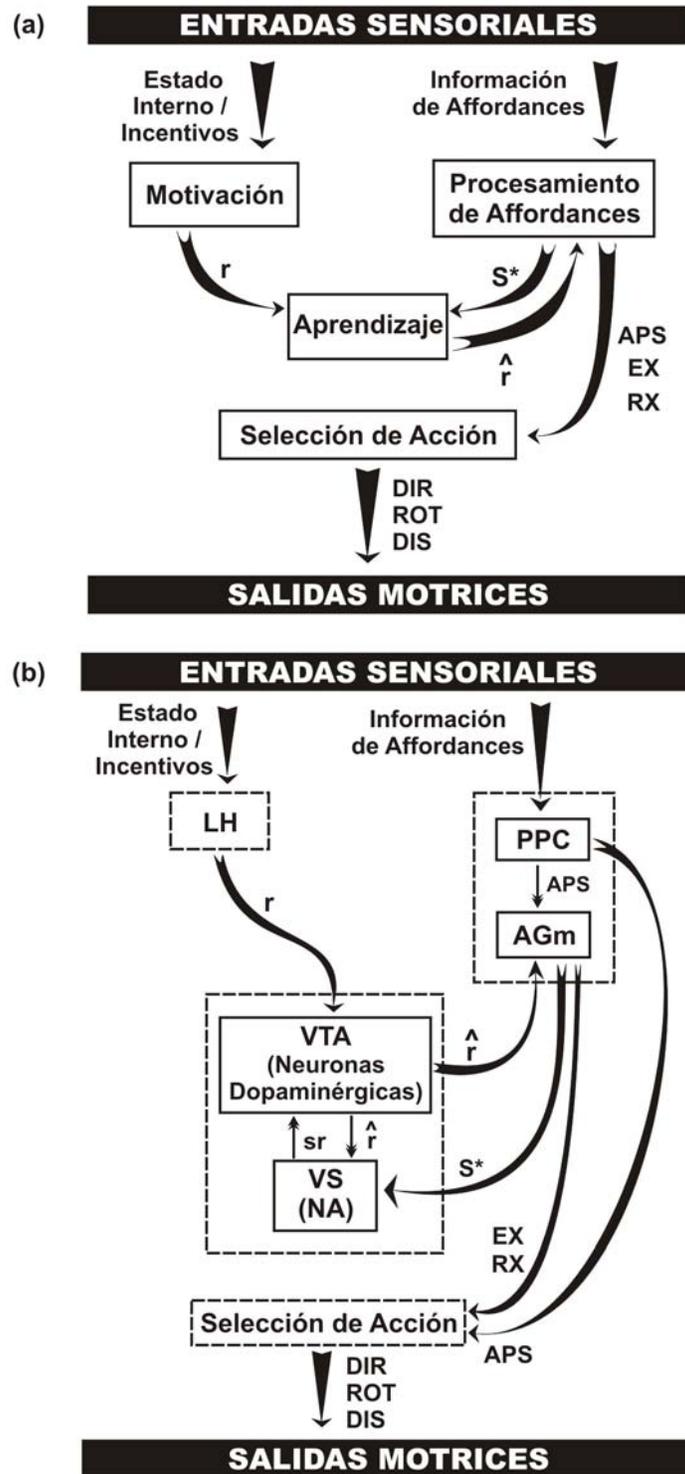


Figura V.2. (a) Un modelo de navegación para el caso de ratas con hipocampo lesionado, y (b) su posible correspondencia biológica. Glosario: r = refuerzo primario; \hat{r} = refuerzo efectivo; S^* = estado actual de affordances; APS= esquema perceptual de affordances; EX= expectativas de recompensa correspondientes a S^* ; RX= rotaciones asociadas a las EX; DIR= próxima dirección de la rata; ROT= rotación de la rata; DIS= próximo desplazamiento del animal.

V.1.2.1. Procesamiento de Affordances

En ausencia de hipocampo, la participación del sistema de mapeo cognitivo en el proceso de aprendizaje se pierde, y éste queda determinado únicamente por información motivacional e información relacionada con las affordances de movimiento. De aquí que se precise mantener la comunicación entre el hipotálamo lateral y el área ventral tegmental de los ganglios basales, y agregar la interacción entre la corteza parietal posterior, a quien se atribuye la generación el esquema perceptual de affordances (APS), y el núcleo estriado.

La corteza parietal posterior (PPC) de la rata es parte de una red cortical que incluye el área agranular medial (AGm) también referida como corteza precentral medial o corteza prefrontal medial, y el área orbital ventrolateral. Se sabe que estas tres regiones corticales juntas están involucradas en el procesamiento espacial y en la orientación del animal, y que PPC se proyecta directamente sobre AGm (Corwin y Reep, 1998). Aún cuando PPC contribuye con entradas al estriado, se ha sugerido que la principal fuente cortical de este núcleo procede de AGm (Cheatwood et al., 2003), como lo ilustra la Figura V.2(b).

La información de movimiento codificada en APS debe ser procesada para contribuir al aprendizaje del animal lesionado. El procesamiento propuesto de APS involucra la introducción de dos capas al modelo que se encargan de generar y detectar patrones de affordances (AFDL), y de determinar si dichos patrones son reconocidos por la rata como previamente experimentados (ASL). Tales funciones son atribuidas a AGm considerando la importante influencia que ejerce sobre el núcleo estriado del cerebro de la rata. La Figura V.3 ilustra la arquitectura del módulo propuesto de procesamiento de affordances.

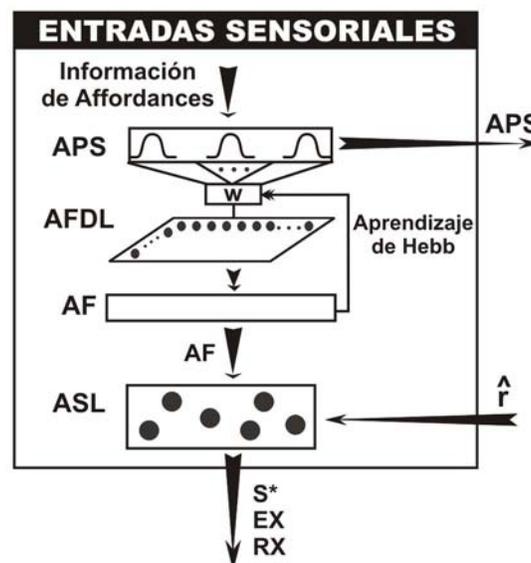


Figura V.3. Módulo de procesamiento de affordances del modelo. Glosario: APS= esquema perceptual de affordances; AFDL= capa detectora de patrones de affordances; AF= patrón de información de affordances; ASL= capa de estados de affordances; \hat{r} = refuerzo efectivo; S^* = estado actual de affordances; EX = expectativas de recompensa correspondientes a S^* ; RX = rotaciones asociadas a las EX .

El esquema APS se proyecta a una capa que genera y detecta patrones de affordances y que es referida como AFDL. Cada neurona en APS es conectada aleatoriamente al 50% de las neuronas en AFDL. Los pesos de conexión w entre las neurona i en APS y j en AFDL son inicializados aleatoriamente y normalizados entre 0 y 1 como lo muestra la ecuación (7).

El nivel de activación de la neurona j en AFDL se calcula sumando los productos entre cada valor de entrada i procedente de la neurona i en APS y el correspondiente peso de conexión w_{ij} como se describe en la ecuación (8).

Se utiliza la regla de aprendizaje de Hebb para actualizar los pesos de conexión entre las capas. La aplicación de esta regla se efectúa de la misma manera que en los módulos de integración de trayectoria, procesamiento de landmarks y representación de lugares del modelo originalmente propuesto. Así, la ecuación (9) permite determinar nuevos valores de activación G_j para las N neuronas j en AFDL, los cuales son almacenados en un arreglo lineal de células referido como AF :

$$AF_j = G_j \mid 1 \leq j \leq N. \quad (28)$$

AF es empleado en la aplicación de la regla de Hebb que se muestra en la ecuación (11), considerando I_i como el valor de entrada procedente de la neurona i en APS. Los pesos actualizados son después normalizados entre 0 y 1 utilizando la ecuación (7). A través de la regla de Hebb, el modelo asegura que la próxima vez que el mismo patrón de activación se presente en APS, se active el mismo conjunto de unidades en AFDL representando un patrón de información de affordances AF producido por AFDL.

Cada patrón de affordances AF es enviado a la capa ASL, encargada de determinar si dicho patrón se reconoce como uno previamente experimentado por la rata, e influir en el proceso de aprendizaje de acciones que conduzcan al animal a alcanzar una recompensa.

Específicamente, cada patrón AF no reconocido por ASL es almacenado en una estructura referida como estado de affordances (S), el cual se asocia con un valor o peso, un rastro de elegibilidad, valores de expectativa de recompensa de cada uno de los giros o rotaciones que comprende, y rastros de elegibilidad correspondientes a tales giros.

Para determinar si la rata reconoce el patrón de affordances actual AF , ASL lo busca dentro de todos los estados de affordances S previamente creados. Esta búsqueda implica el cálculo del grado de similitud SD entre AF y cada patrón almacenado pat , como se describe en la ecuación (15) para el caso de reconocimiento de lugares (Sección III.6).

Al respecto, el modelo distingue entre dos casos: (i) si existe al menos un valor SD que exceda cierto valor de umbral cercano a 1, el estado S que almacena el patrón de affordances con el mayor valor SD es considerado como el ganador y se convierte en el estado activo; (ii) si no existe un S ganador, ASL crea un nuevo estado para almacenar el patrón actual AF y éste se convierte en el estado activo.

Para contribuir a la selección de la próxima acción a ejecutar por la rata, ASL entrega al módulo de selección de acción del modelo los valores de expectativa de recompensa (EX) asociados a las diversas rotaciones (RX) incluidas en el estado de affordances activo S^* . Tales expectativas son actualizadas en ASL a través del módulo de aprendizaje que se describe a continuación.

V.1.2.2. Aprendizaje

Como en el modelo original, se emplea una arquitectura de aprendizaje por refuerzo Actor-Crítico con el fin de procesar valores esperados de recompensas futuras a través de un Crítico Adaptivo (AC) y un número de unidades Actor. La Figura V.4 ilustra la estructura de los módulos involucrados en el proceso de aprendizaje.

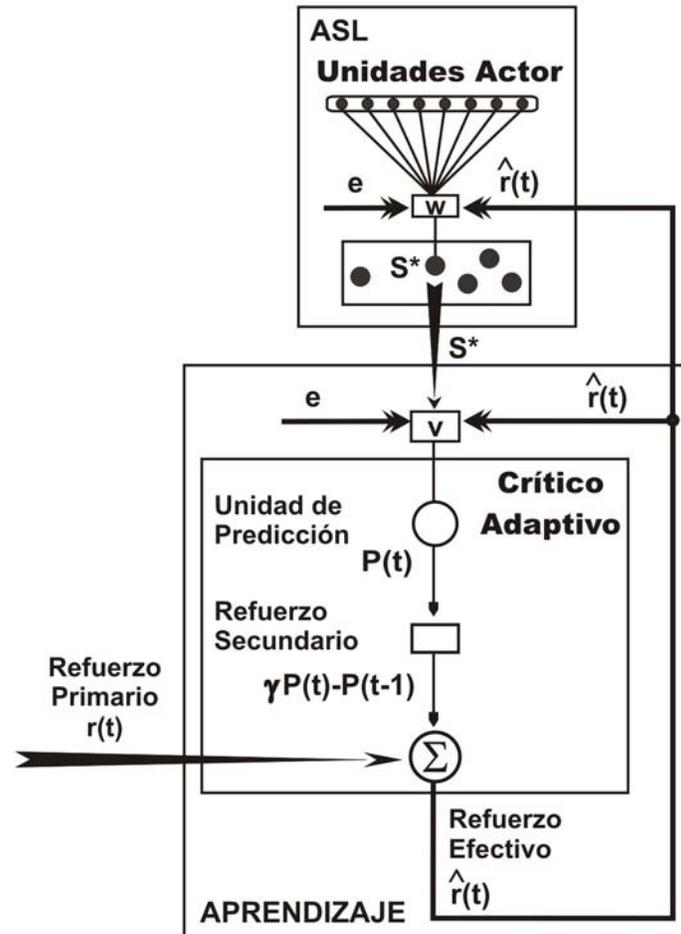


Figura V.4. El módulo de aprendizaje del modelo. ASL= capa de estados de affordances; e = rastros de elegibilidad; \hat{r} = refuerzo efectivo; w = pesos de conexión entre un estado de affordances y sus unidades Actor; S^* = estado de affordances activo; v = peso global de S^* ; $P(t)$ = valor estimado de futura recompensa para S^* en el tiempo t .

AC incluye una unidad de predicción que estima en cada instante de tiempo t el valor de recompensa futura P del estado de affordances activo S^* . Dicho valor corresponde al peso global v del estado S^* :

$$P(t) = v(t). \quad (29)$$

Como en el módulo de aprendizaje originalmente propuesto, AC utiliza las predicciones calculadas en los tiempos t y $t-1$ para determinar el refuerzo secundario descontando la predicción actual a una tasa γ para obtener su valor presente. La suma del refuerzo secundario y el refuerzo primario r calculado por el módulo de motivación del modelo constituye el refuerzo efectivo \hat{r} como se describe en la ecuación (18) de la Sección III.7.

Las diferentes rotaciones comprendidas dentro de S^* son consideradas como unidades Actor. Así, cada estado de affordances puede estar conectado a un número máximo de ocho unidades Actor, y cada conexión está asociada con un peso w y un rastros de elegibilidad e . Los ocho posibles pesos representan la expectativa del animal por obtener una recompensa al ejecutar la rotación correspondiente a la unidad Actor específica, mientras que los rastros de elegibilidad permiten marcar las conexiones para reforzarlas en un tiempo posterior.

El refuerzo efectivo \hat{r} es empleado para actualizar los pesos globales v asociados a los estados de affordances, así como los pesos w correspondientes a cada una de las rotaciones comprendidas dentro de los estados de affordances. En el primer caso, el modelo utiliza (30):

$$v_S(t+1) = v_S(t) + \beta \hat{r}(t) e_S(t), \quad (30)$$

donde $v_S(t+1)$ es el peso global del estado de affordances S en el tiempo $t+1$, β es la tasa de aprendizaje, y $e_S(t)$ es el rastro de elegibilidad del mismo estado S en el tiempo t .

En el segundo caso se utiliza la ecuación (31):

$$w_S^d(t+1) = w_S^d(t) + \beta \hat{r}(t) e_S^d(t), \quad (31)$$

donde $w_S^d(t+1)$ es el peso o expectativa de recompensa del estado S que corresponde a la rotación d en el tiempo $t+1$, y $e_S^d(t)$ es el rastro de elegibilidad de la rotación d dentro del estado S en el tiempo t .

Como se muestra en (30) y (31), ambas reglas de aprendizaje dependen de la elegibilidad de los estados o de las rotaciones dentro de los estados, respectivamente. Al inicio de cualquier ensayo en un experimento dado, todos los rastros de elegibilidad son inicializados en 0. En cada instante de tiempo t dentro de un ensayo, el rastro de elegibilidad e del estado activo S^* se incrementa si la acción ejecutada por la rata en el tiempo $t-1$ le permitió percibir la meta, como se muestra en la ecuación (32):

$$e_{S^*}(t) = e_{S^*}(t-1) + \chi, \quad (32)$$

donde χ es el parámetro de incremento.

Resultados derivados de la experimentación con ratas han demostrado que el daño en el hipocampo no tiene efecto en la memoria inmediata o de corto plazo, pues la corteza prefrontal comparte dicha función con el hipocampo (Lee y Kesner, 2003). Sin embargo, el hipocampo resulta una estructura esencial cuando se trata de recordar eventos no inmediatos (i.e., que ocurrieron hace más de 5 minutos). Específicamente, en el laberinto radial de ocho brazos, las ratas sanas son capaces de recordar, en el orden correcto, los primeros dos brazos elegidos y los últimos dos, en tanto que las ratas con hipocampo lesionado sólo pueden recordar las últimas dos decisiones inmediatas y el orden en que ocurrieron (Kesner y Novak, 1982).

El modelo considera esta evidencia para reforzar no sólo la última decisión de la rata, sino también la penúltima, así que al inicio de cualquier ensayo del experimento, se actualiza el rastro de elegibilidad e del estado de affordances S_p que corresponde a la localidad en que el animal tomó su penúltima decisión como lo indica la ecuación (33):

$$e_{S_p}(t) = e_{S_p}(t-1) + \chi. \quad (33)$$

En este caso, si la rata llegó a la meta al final del ensayo anterior, el rastro de elegibilidad e del estado S_p se incrementa (i.e., se utiliza χ positivo), mientras que si no llegó a ella, se decrementa (i.e., se utiliza χ negativo).

Por otro lado, en cada instante de tiempo t dentro de un ensayo, el rastro de elegibilidad e que corresponde a la rotación d ejecutada por la rata en el tiempo $t-1$ dentro del estado activo S^* se incrementa si tal acción le permitió percibir la meta, como se muestra en la ecuación (34):

$$e_{S^*}^d(t) = e_{S^*}^d(t-1) + \tau. \quad (34)$$

donde τ es el parámetro de incremento. Al inicio de cualquier ensayo del experimento, el modelo actualiza también el rastro de elegibilidad e del estado S_p que corresponde a la rotación d ejecutada en el lugar en que el animal tomó su penúltima decisión, como lo indica (35):

$$e_{S_p}^d(t) = e_{S_p}^d(t-1) + \tau. \quad (35)$$

En este caso, si la rata llegó a la meta al final del ensayo anterior, se utiliza τ positiva; en caso contrario, τ negativa. Adicionalmente, considerando que las ratas lesionadas sólo pueden utilizar una hipótesis de orientación durante el aprendizaje de tareas espaciales, la frustración ocasionada por la extinción de una recompensa dispara su curiosidad por las posibles orientaciones adyacentes a la actual. De esta forma, al inicio de cualquier ensayo, el modelo incrementa el rastro de elegibilidad e del estado S_p que corresponde a la rotación $d+1$ adyacente a la rotación d , la cual fue ejecutada en el lugar en que el animal tomó su penúltima decisión y le provocó frustración:

$$e_{S_p}^{d+1}(t) = e_{S_p}^{d+1}(t-1) + \tau. \quad (36)$$

Finalmente, después de actualizar los pesos v asociados a los diversos estados de affordances y las expectativas de recompensa w que corresponden a las rotaciones d comprendidas dentro de cada estado de affordances S en cualquier instante de tiempo t dentro de un ensayo, todos los rastros de elegibilidad e decaen a una cierta tasa λ , como se muestra en la ecuación (37):

$$\begin{aligned} e_S(t) &= \lambda e_S(t-1) \\ e_S^d(t) &= \lambda e_S^d(t-1) \end{aligned} \quad (37)$$

V.1.2.3. Selección de Acción

Para determinar la próxima dirección del animal lesionado en cualquier instante de tiempo t , el módulo de selección de acción considera tres señales codificadas a través de esquemas perceptuales que corresponden a

- las affordances disponibles en el tiempo t (APS),
- una rotación aleatoria entre las affordances disponibles en el tiempo t (RPS), y
- la expectativa global de máxima recompensa (EMR).

En esta versión adaptada del módulo de selección de acción, el esquema perceptual EMR es construido a partir de los valores de los pesos o expectativas de recompensa EX entregadas por la capa ASL del modelo. La cantidad de valores EX corresponde con la cantidad de posibles rotaciones incluidas en el estado actual de affordances S^* . Cada valor de expectativa de recompensa EX_k es representado como una distribución de Gauss en EMR centrada en la posición del arreglo correspondiente a la rotación RX_k con la que está asociado, y cuya altura depende del valor EX_k . De esta forma, EMR se inicializa de la misma forma descrita en la ecuación (24) para el caso de la versión original del módulo (Sección III.8), donde M_k se define como $4+9m$ con m un valor entero entre 0 y 8 que en este caso corresponde a la rotación RX_k .

Para generar una señal de expectativa global de recompensa que influya finalmente en el próximo comportamiento de la rata, el módulo de selección de acción calcula el "centro de masa" c sobre EMR mediante la misma ecuación (25) presentada en la Sección III.8.

Posteriormente, EMR es actualizado para representar la expectativa global de recompensa por medio de una sola curva de Gauss centrada en c con altura equivalente a la suma de las alturas de cada campana k preexistente en EMR. La ecuación (26) describe dicha actualización. En caso de no existir una affordance disponible que coincida con el centro de masa en EMR, dicho centro es desplazado a la posición del arreglo asociada con la affordance más cercana.

Finalmente, el módulo de selección de acción suma, neurona por neurona, los valores de activación almacenados en los esquemas perceptuales APS, RPS y EMR produciendo un nuevo esquema perceptual O , donde el valor de activación de cualquiera neurona i se calcula como se describe enseguida:

$$O_i = APS_i + RPS_i + EMR_i . \quad (38)$$

La influencia de cada señal en O depende de la altura de sus correspondientes distribuciones Gaussianas. Específicamente, el orden de importancia de tales señales en la selección de la próxima acción a ejecutar por la rata es el siguiente: (i) EMR, (ii) APS, y (iii) RPS.

En el arreglo resultante O , el módulo de selección de acción considera la posición de la neurona con el valor de activación más alto para determinar la próxima dirección de la rata DIR , entre 0° y 315° en intervalos de 45° , y la rotación ROT requerida para apuntar hacia tal dirección. Si DIR es diferente de la dirección actual, el próximo desplazamiento de la rata DIS es 0, dando al animal la oportunidad de obtener una perspectiva diferente desde el mismo lugar. De otra forma, DIS es 1, valor que corresponde a “un paso” en dirección DIR .

El valor de los principales parámetros involucrados en esta versión del modelo para ratas lesionadas se presenta en el Apéndice A.

V.1.3. Resultados Experimentales

Por simplicidad en la implementación de la tarea de discriminación inversa de O’Keefe utilizando un robot, el modelo original y el modelo adaptado para animales lesionados fueron probados en el laberinto tipo T y en el radial de 8 brazos separadamente, en lugar de intercalar ensayos de prueba en ambos laberintos. Adicionalmente, la misma tarea espacial fue implementada en un laberinto extendido compuesto por múltiples Ts utilizando la versión del modelo originalmente propuesta. Los siguientes apartados describen los resultados obtenidos en cada caso.

V.1.3.1. El Laberinto Tipo T y el Radial de 8 brazos con Ratas Sanas

En el laberinto tipo T mostrado en la Figura V.5(a), el robot navega desde la base de la T hasta el fondo de cualquiera de los brazos, y después regresa a su localidad de partida. Este proceso se repite en cada ensayo del experimento. Es importante señalar que el regreso autónomo no es parte del protocolo originalmente implementado por O’Keefe. De hecho, al término de cualquier ensayo dado, las ratas eran manualmente colocadas de nuevo en la base de la T. No obstante, decidimos automatizar el proceso de retorno con el fin de ejecutar el experimento completo sin intervención humana, excepto para cambiar la localización de la meta al inicio de las pruebas.

Durante la fase de entrenamiento, la meta es colocada al fondo del brazo izquierdo del laberinto y el robot es capaz de reconocerla sólo a un paso de distancia. Al inicio del entrenamiento, las decisiones del robot en la localidad de bifurcación de la T están determinadas por la curiosidad de ejecutar nuevas rotaciones y por ruido, el cual es representado por el módulo de selección de acción del modelo como una rotación aleatoria. Después de haber explorado cada brazo una vez, el nivel de curiosidad por rotar -90° y $+90^\circ$ en el punto de bifurcación decrece, prevaleciendo el ruido. Después, el robot gira a la izquierda o a la derecha de manera

aleatoria, y eventualmente alcanza el criterio una vez que la expectativa por encontrar una recompensa al orientarse a 180° en el punto de bifurcación se vuelve mayor que el ruido. Cuando este evento ocurre, la fase de entrenamiento concluye.

El robot ejecuta tantos ensayos como requiere para aprender a dirigirse al brazo izquierdo y obtener así la recompensa. La duración promedio de la fase de entrenamiento en términos de la cantidad de ensayos fue 12, valor obtenido al ejecutar seis veces la versión original del modelo. Considerando que un ensayo dura 2 minutos en promedio, el robot completó la fase de entrenamiento en poco más de 20 minutos.

Cuando la fase de prueba inicia, la meta es cambiada al fondo del brazo derecho de la T. Durante la tarea inversa, la expectativa de futura recompensa por el brazo izquierdo decrece continuamente cada vez que el robot llega al final del corredor y no encuentra la recompensa, dado que este evento es codificado como frustrante por la regla de aprendizaje por refuerzo del modelo. Cuando la expectativa de recompensa se vuelve más pequeña que el ruido, el robot comienza a visitar el brazo derecho incrementando así la expectativa de recompensa por esta opción. Mientras el valor de expectativa es menor que el ruido, el robot tiende a elegir cualquier brazo de manera aleatoria hasta que alcanza el criterio cuando la expectativa de recompensa por girar a la derecha en el punto de bifurcación es más grande que el ruido. El desempeño promedio del robot durante la tarea de discriminación inversa derivado de seis ejecuciones del modelo es mostrado en la gráfica superior de la Figura V.6(a), donde el comportamiento del robot está expresado en términos del porcentaje de elecciones correctas en la localidad de bifurcación de la T.

Como resultado del entrenamiento, el 100% de las decisiones del robot fueron correctas (control). La gráfica muestra 32 ensayos de prueba. Como se ve, el robot tarda 12 ensayos en desaprender la opción previamente correcta (criterio), y a partir del ensayo 13 ha aprendido la nueva. De esta forma, el porcentaje de elecciones correctas cambia de 36% en el ensayo 12 a 95% en el ensayo 16 y a 100% en adelante.

Comparando los resultados anteriores con los reportados por O'Keefe con ratas sanas en (O'Keefe, 1983), es posible apreciar la similitud conductual con el robot en el laberinto tipo T. La gráfica superior de la Figura V.6(b) ilustra los resultados promedio de cuatro ratas presentados por O'Keefe. Esta gráfica también muestra 32 ensayos de prueba y una medida de control del 100% de elecciones correctas en la localidad de bifurcación después del entrenamiento. En este caso, las ratas alcanzaron el criterio en el ensayo 20, donde el porcentaje de elecciones correctas fue un valor entre 20% y 40%. Para el ensayo 24, las ratas eligieron el nuevo brazo con recompensa en más del 90% de las veces hasta llegar al 100% de las decisiones correctas. Tanto en la gráfica superior de la Figura V.6(a) como en la de la Figura V. 6(b) reportada por O'Keefe, existe un cambio abrupto del brazo incorrecto al brazo correcto.

Durante el experimento, el robot construye y mantiene la representación espacial topológica-métrica mostrada en la Figura V.5(b). En esta tarea de discriminación, la actividad conjunta de la población de células de lugar fue determinada a partir de información kinestésica únicamente, dado que no se dispusieron landmarks alrededor del laberinto. La Figura V.5(c) ilustra una muestra de la actividad conjunta registrada por la capa de células de lugar (PCL) del modelo cuando el robot alcanzó la localidad de bifurcación de la T indicada con la etiqueta "e" en la Figura V.5(a) estando orientado a 90° . Aún cuando el 25% de las células de lugar mostraron campos receptivos hasta cierto grado en esta localidad, la figura sólo muestra las cinco neuronas más activas encontradas dentro de cada vecindario en que la capa fue dividida. Los campos de lugar traslapados del 25% de las células son mapeados al nodo 3 dentro de la representación espacial, donde el patrón de actividad de la colección de células de lugar es almacenado en una unidad Actor asociada con la dirección 90° .

A diferencia de lo que ocurre con las ratas, las células de lugar en el modelo propuesto responden de manera unidireccional; esto es, cuando el robot gira a la izquierda en el punto de bifurcación de la T, otro grupo de neuronas responde. En esta situación, la actividad global actual de la capa es almacenada en una nueva unidad Actor asociada con la dirección 180° y conectada al mismo nodo 3 del mapa.

La relevancia de las localidades en un laberinto depende de la presencia de una recompensa o de los cambios de affordances percibidos por el robot durante la exploración. En la Figura V.5(a), por ejemplo, el

robot considera relevantes las localidades “a”, “b”, “e”, “f”, “g”, “h” e “i”, por lo que el mapa incluye siete nodos. Cuando el robot llega a la localidad “b” en dirección 90° , la actividad conjunta actual de las neuronas en PCL es almacenada en el nodo 2, y aunque el patrón de actividad pudiera variar ligeramente en las localidades “c” y “d”, las affordances percibidas por el robot no cambiaron de “b” a “c” o “d”, por lo que los patrones de actividad registrados en estas localidades son promediados y almacenados en el mismo nodo 2, definiendo de esta forma su extensión física.

Al terminar cualquier ensayo dado, el robot regresa a la localidad de partida leyendo las direcciones almacenadas en el mapa y no por decisiones aleatorias, así que el mapa no es modificado con nuevos arcos durante este procedimiento. El proceso de retorno seguido por el robot en este laberinto tipo T fue documentado de manera preliminar en (Barrera y Weitzenfeld, 2006).

Finalmente conviene hacer notar que, a diferencia de las ratas, el robot fue programado para evitar rotaciones de 180° cuando existen otras opciones (0° , $+90^\circ$, -90°) con el fin de optimizar el proceso de exploración para encontrar la meta. Por esta razón, las direcciones de los arcos que conectan los nodos del mapa son mostradas en un solo sentido.

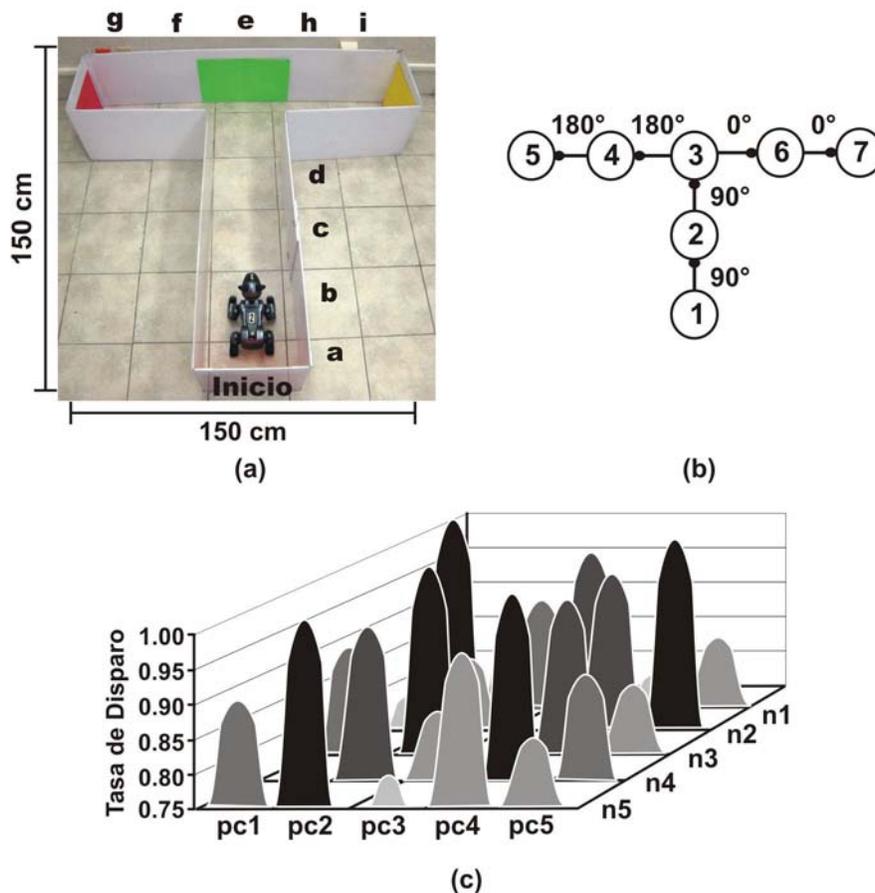


Figura V.5. (a) El laberinto tipo T utilizado en la tarea de discriminación inversa. Las diversas localidades están etiquetadas con letras. El robot Sony AIBO está localizado en la posición de inicio. (b) El mapa construido por el robot durante el entrenamiento. Los nodos se presentan numerados en orden de creación, y los arcos están asociados con la dirección del robot cuando se movió de un nodo al siguiente. (c) Una muestra de la actividad conjunta de las células de lugar cuando el robot alcanza la localidad de bifurcación estando orientado a 90° . La figura muestra sólo la tasa de disparo de las cinco neuronas más activas (pc1, pc2, pc3, pc4, pc5) de cada vecindario (n1, n2, n3, n4, n5) dentro de la capa de células de lugar del modelo.

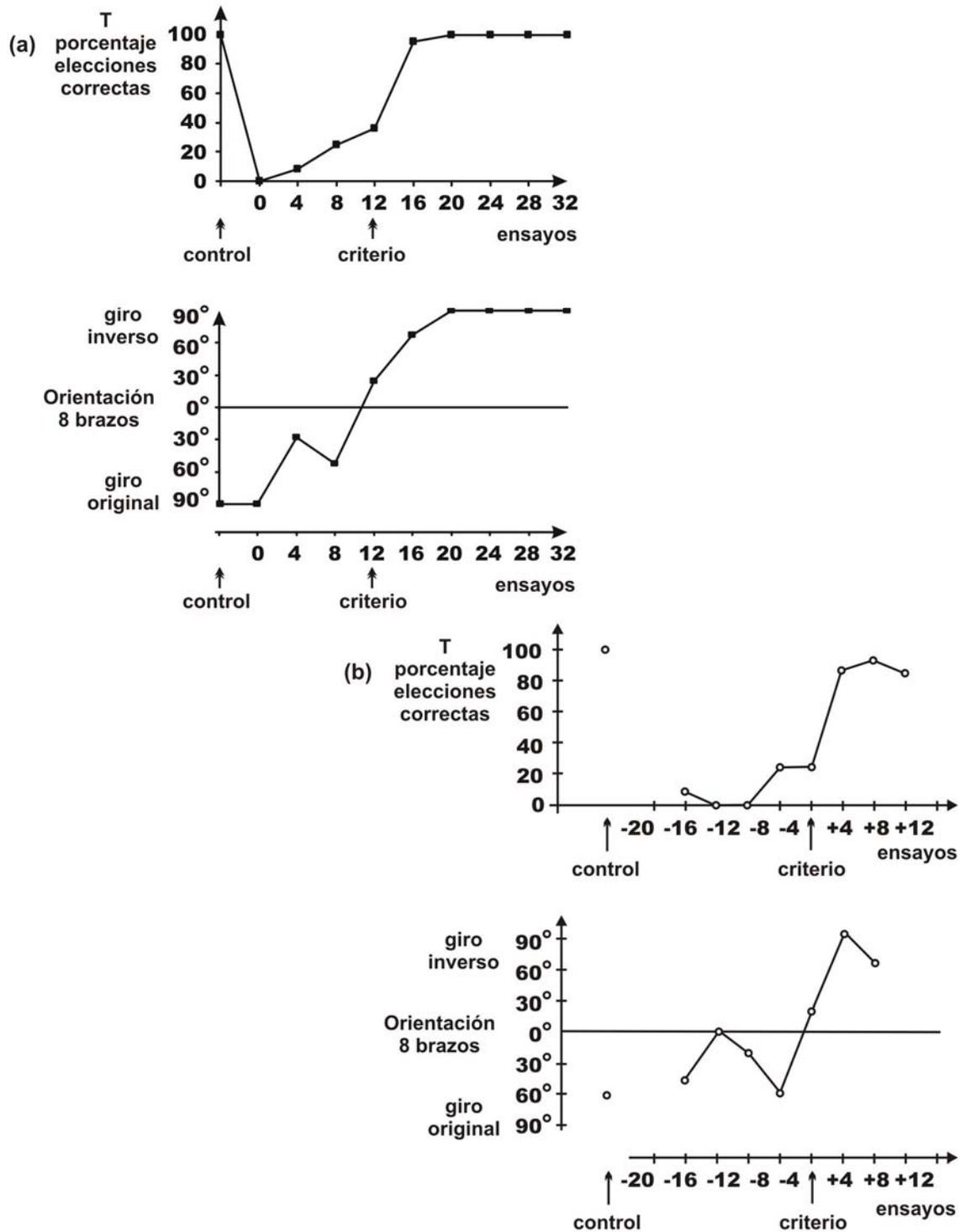


Figura V.6. (a) El desempeño de seis robots durante la tarea de discriminación inversa de O'Keefe. Cada gráfica fue obtenida promediando las gráficas de los robots individuales. La gráfica superior muestra el porcentaje de elecciones correctas en el laberinto tipo T promediadas en periodos de cuatro ensayos. La gráfica inferior presenta las decisiones del robot en el laberinto radial también promediadas en periodos de cuatro ensayos. (b) El desempeño de cuatro ratas sanas durante la tarea de discriminación inversa reportado por O'Keefe y extraído de (O'Keefe, 1983). Cada gráfica fue obtenida promediando las gráficas de los animales individuales. En (a) y en (b), el cambio abrupto de girar a la izquierda a girar a la derecha en el laberinto T revela el momento (criterio) en que la orientación promedio cruza la línea media en el laberinto radial de 8 brazos.

La misma tarea de discriminación inversa fue implementada en el laberinto radial de 8 brazos presentado en la Figura V.7(a). En cualquier ensayo dado, el robot inicia en la localidad indicada en la figura, navega al fondo de cualquiera de los otros brazos, y regresa al punto de partida. La Figura V.7(b) muestra el mapa construido por el robot durante el experimento.

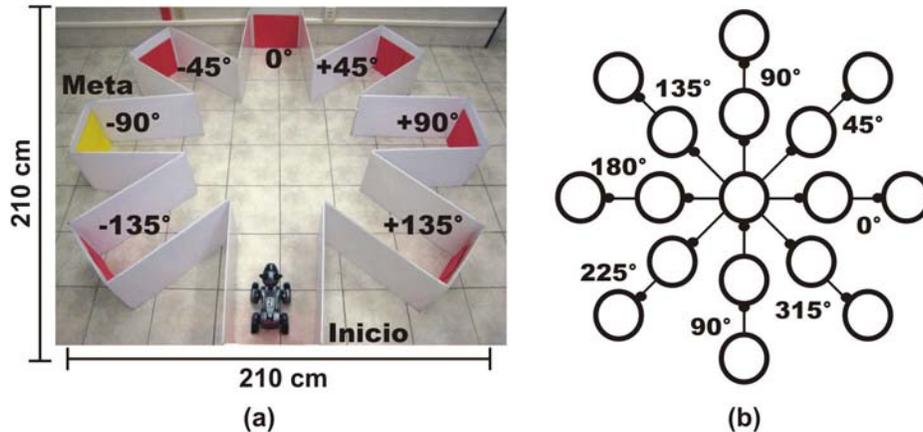


Figura V.7. (a) El laberinto radial de 8 brazos utilizado durante la tarea de discriminación inversa de O'Keefe. La fotografía muestra al robot Sony AIBO localizado en el punto de partida, así como las direcciones egocéntricas de los brazos del laberinto con respecto a dicha localidad. (b) El mapa construido por el robot durante el experimento. Los arcos entre nodos están asociados con la dirección allocéntrica del robot cuando se movió de un nodo al siguiente.

La fase de entrenamiento funciona como el laberinto T con la meta colocada al fondo del brazo relativamente orientado a -90° . Al inicio, las acciones del robot en la localidad de decisión cuando llega al centro del laberinto están determinadas por curiosidad y ruido. Después de visitar cada brazo una vez, el nivel de curiosidad decrece, prevaleciendo ruido. Eventualmente, el robot alcanza el criterio una vez que la expectativa de encontrar una recompensa al orientarse a 180° en la localidad de decisión se vuelve más grande que el ruido. Cuando este evento ocurre, la fase de entrenamiento concluye.

La duración promedio de la fase de entrenamiento obtenida después de ejecutar el modelo seis veces fue de 13 ensayos completados por el robot en menos de media hora.

Durante la tarea inversa, la expectativa de futura recompensa por el brazo orientado a -90° decrece continuamente cada vez que el robot llega al fondo del corredor y no encuentra la recompensa. Cuando la expectativa de recompensa se vuelve menor que el ruido, el robot comienza a visitar otros brazos aleatoriamente. Cada vez que visita el brazo orientado a $+90^\circ$ que ahora provee la recompensa, la expectativa por este corredor aumenta. El robot alcanza el criterio cuando la expectativa de recompensa por girar a la derecha en la localidad de decisión es mayor que el ruido. El desempeño promedio del robot durante las pruebas en la tarea inversa que resultó de seis ejecuciones del modelo es mostrado en la gráfica inferior de la Figura V.6(a), la cual presenta las elecciones del robot durante 32 ensayos agrupados de cuatro en cuatro.

Como resultado del entrenamiento, el robot eligió consistentemente girar a la izquierda en la localidad central del laberinto radial (control). Durante las pruebas, la orientación del robot no reveló algún cambio sistemático. Como en el laberinto T, el criterio ocurrió alrededor del ensayo 12, cuando la orientación promedio cruza la línea media de la gráfica.

En este caso, nuestros resultados también muestran una similitud con los resultados reportados por O'Keefe con cuatro ratas sanas en el laberinto radial (O'Keefe, 1983). La gráfica inferior de la Figura V.6(b) presenta los resultados de O'Keefe. Se muestran 32 ensayos de prueba, donde, en promedio, las ratas alcanzaron el criterio alrededor del ensayo 20. Él explicó que el cambio abrupto de girar a la izquierda a girar a la derecha en el laberinto T es un reflejo del momento en que la orientación promedio cruza la línea media

en el laberinto radial. De acuerdo a nuestras pruebas, el mismo hecho aplica para el comportamiento del robot.

V.1.3.2. El Laberinto Tipo T y el Radial de 8 brazos con Ratas Lesionadas

El modelo para ratas con hipocampo lesionado fue empleado para reproducir con robots la tarea de discriminación inversa de O'Keefe. Al igual que en el caso de las ratas sanas, el experimento fue implementado con el robot en el laberinto T (Figura V.5(a)) y en el radial de 8 brazos (Figura V.7(a)) separadamente.

El protocolo de entrenamiento en el laberinto T fue el siguiente: en cada ensayo, el robot navega de la base de la T al fondo de cualquiera de los corredores horizontales, y al terminar el recorrido, el robot es colocado manualmente en la base para iniciar el siguiente ensayo. Dado que el robot con "hipocampo lesionado" no tiene la capacidad de construir un mapa del laberinto, no puede regresar al punto de partida de manera autónoma.

Durante el entrenamiento, la meta se encuentra localizada al fondo del extremo izquierdo de la T. El robot ejecuta tantos ensayos como requiere para alcanzar el criterio una vez que la expectativa por encontrar una recompensa al rotar -90° en el punto de bifurcación y avanzar hasta el final del corredor se vuelve mayor que el ruido. Cuando este evento ocurre, la fase de entrenamiento concluye pues el robot ha aprendido la tarea. Considerando que el modelo fue ejecutado seis veces, la duración promedio del entrenamiento fue de 11 ensayos, y la medida de control resultante consistió en el 100% de las decisiones correctas; es decir, el robot logró girar consistentemente a la izquierda al llegar al punto de elección del laberinto.

Durante las pruebas, la meta es colocada al fondo del pasillo derecho de la T. La frustración que obtiene el robot cada vez que se orienta a -90° en el punto de bifurcación de la T y llega al fondo del corredor, provoca un decremento en su expectativa por encontrar ahí una recompensa y un incremento en su expectativa por la orientación adyacente $+90^\circ$ dentro del estado de affordances [-90° , $+90^\circ$ y 180°]. Cuando el valor de expectativa de recompensa por girar -90° en el punto de bifurcación de la T llega a ser menor que el ruido, el robot comienza a ejecutar la rotación $+90^\circ$, lo cual implica un incremento adicional en su expectativa por esta nueva opción recompensada. Eventualmente, cuando este valor de expectativa se vuelve mayor al ruido, el robot alcanza el criterio. El desempeño promedio del robot "lesionado" durante la tarea de discriminación inversa derivado de seis ejecuciones del modelo es mostrado en la gráfica superior de la Figura V.8(a), donde el comportamiento del robot está expresado en términos del porcentaje de elecciones correctas en la localidad de bifurcación de la T.

La gráfica muestra 32 ensayos de prueba. Como se ve, el robot "lesionado" tarda en promedio 12 ensayos en desaprender la opción previamente correcta (criterio), y a partir del ensayo 13 ha aprendido a rotar a la derecha. De esta forma, el porcentaje de elecciones correctas pasa de 35% en el ensayo 12 a 95% en el ensayo 16 y a 100% en adelante.

Es posible apreciar la similitud conductual entre los resultados obtenidos con el robot en el laberinto tipo T, y los resultados derivados por O'Keefe al experimentar con ratas lesionadas (O'Keefe, 1983). La gráfica superior de la Figura V.8(b) ilustra el desempeño promedio de seis ratas lesionadas presentado por O'Keefe. Esta gráfica también muestra 32 ensayos de prueba y una medida de control del 100% de elecciones correctas en la localidad de bifurcación después del entrenamiento. En este caso, las ratas alcanzaron el criterio en el ensayo 20, donde el porcentaje de elecciones correctas fue un valor entre 20% y 40%. Para el ensayo 24, las ratas eligieron el nuevo giro con recompensa en más del 90% de las veces hasta llegar al 100% de las decisiones correctas. Tanto en la gráfica superior de la Figura V.8(a) como en la de la Figura V.8(b) reportada por O'Keefe, existe un cambio abrupto del giro originalmente recompensado al nuevo giro que ofrece una recompensa.

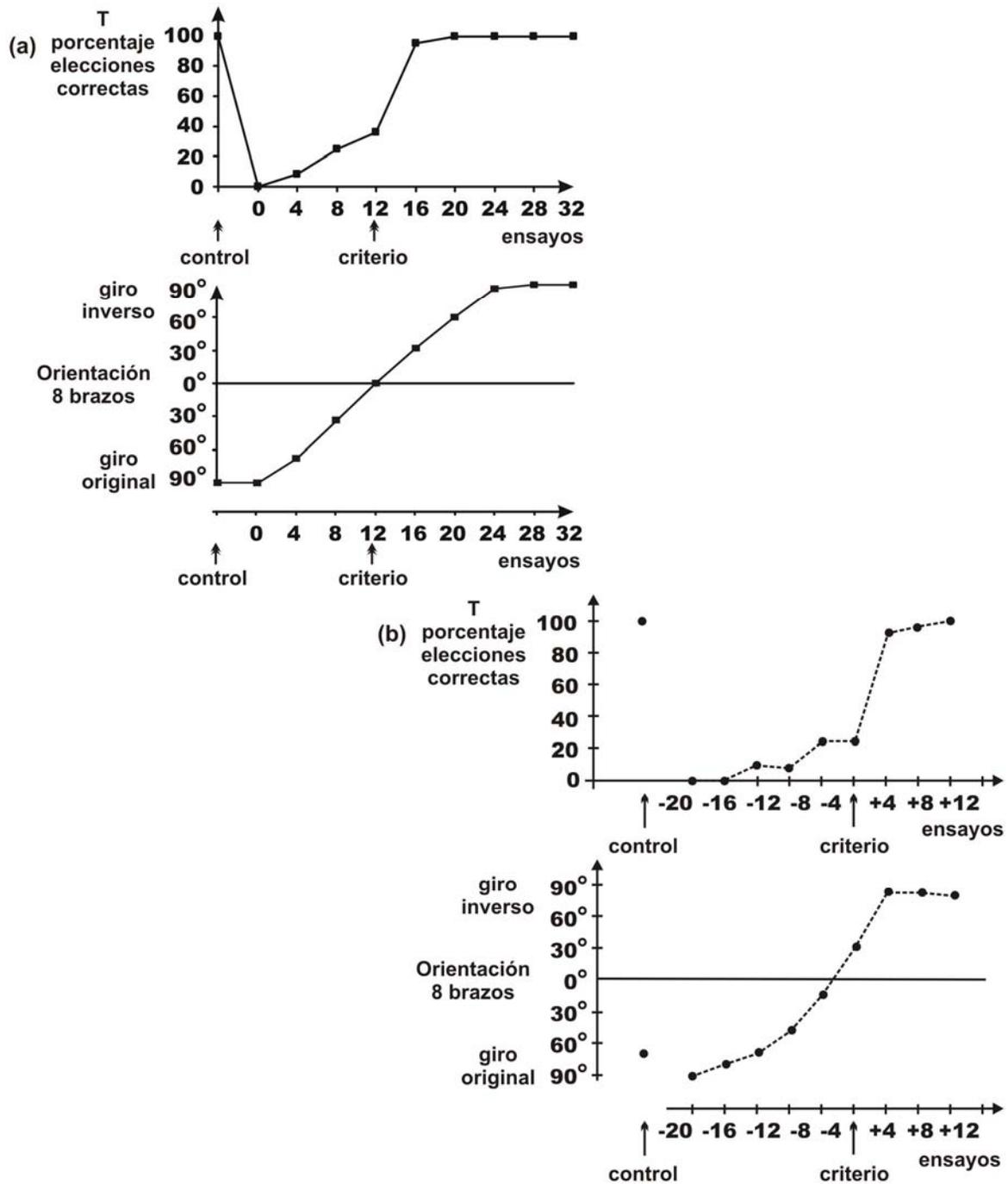


Figura V.8. (a) El desempeño promedio de seis robots "lesionados" durante la tarea de discriminación inversa de O'Keefe. Cada gráfica fue obtenida promediando las gráficas de los robots individuales. La gráfica superior muestra el porcentaje de elecciones correctas en el laberinto tipo T promediadas en periodos de cuatro ensayos. La gráfica inferior presenta las decisiones del robot en el laberinto radial también promediadas en periodos de cuatro ensayos. (b) El desempeño promedio de seis ratas lesionadas durante la tarea de discriminación inversa reportado por O'Keefe y extraído de (O'Keefe, 1983). Cada gráfica fue obtenida promediando las gráficas de los animales individuales. En (a) y en (b) es factible apreciar un cambio abrupto de girar a la izquierda a girar a la derecha en el laberinto T, y un corrimiento sistemático y gradual en la orientación de los robots y de las ratas desde el giro originalmente recompensado (-90°), pasando por seguir derecho (0°) y alcanzando el nuevo giro con recompensa ($+90^\circ$) en el laberinto radial de 8 brazos. Dicho cambio en la T revela el momento (criterio) en que la orientación promedio cruza la línea media en el laberinto radial.

Por otro lado, el protocolo de entrenamiento en el laberinto radial mostrado en la Figura V.7(a) incluyó la navegación del robot del punto de partida al fondo de cualquiera de los 7 corredores orientados relativamente a 0° , $\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$, y al terminar el recorrido, la colocación manual del mismo en el punto de partida para iniciar el siguiente ensayo.

Durante el entrenamiento, la meta se encuentra localizada al fondo del brazo relativamente orientado a -90° . Al inicio, las acciones del robot al llegar al centro del laberinto están determinadas por ruido hasta que, eventualmente, el robot alcanza el criterio una vez que la expectativa de recompensa por girar -90° se vuelve más grande que el ruido. Cuando este evento ocurre, la fase de entrenamiento concluye. La duración promedio de la fase de entrenamiento resultante de ejecutar el modelo seis veces fue de 13 ensayos.

Durante la tarea inversa, la meta se encuentra localizada al fondo del pasillo orientado relativamente a $+90^\circ$. La frustración que obtiene el robot cada vez que se orienta a -90° en el punto central del laberinto y llega al fondo del corredor, provoca un decremento en su expectativa por encontrar ahí una recompensa y un incremento en su expectativa por la orientación adyacente -45° dentro del estado de affordances [-135° , -90° , -45° , 0° , $+45^\circ$, $+90^\circ$, $+135^\circ$]. Cuando el valor de expectativa de recompensa por girar -90° en el punto central llega a ser menor que la expectativa por rotar -45° , el robot comienza a ejecutar esta rotación. El proceso de actualización de expectativas de recompensa por las orientaciones adyacentes continúa hasta que el robot alcanza el criterio cuando la expectativa de recompensa por girar $+90^\circ$ resulta mayor que la expectativa por ejecutar cualquier otra rotación. Como lo muestra la Tabla A.II dentro del Apéndice A, el nivel de ruido (altura de la distribución de Gauss en el esquema perceptual de aleatoriedad RPS) utilizado durante la tarea inversa en el laberinto radial es significativamente inferior al empleado durante el entrenamiento. Este hecho tiene el propósito de determinar los procesos de desaprendizaje y aprendizaje principalmente a través de la expectativa de recompensa del robot, la cual resulta suficiente para permitirle manifestar el comportamiento cualitativo esperado.

El desempeño promedio del robot "lesionado" durante las pruebas que se derivó de ejecutar seis veces el modelo es mostrado en la gráfica inferior de la Figura V.8(a), la cual presenta las elecciones del robot durante 32 ensayos agrupados de cuatro en cuatro. Como resultado del entrenamiento, el robot eligió consistentemente girar a -90° al llegar a la localidad central del laberinto radial (control). Durante las pruebas, el robot tarda en promedio 13 ensayos en desaprender el giro a -90° (criterio), durante los cuales la orientación relativa del robot cambió de -90° a -45° y de -45° a 0° en este orden. En el ensayo 14, su orientación pasa al cuadrante de las rotaciones positivas cambiando de 0° a $+45^\circ$ y de $+45^\circ$ a $+90^\circ$, y permaneciendo ahí hasta el fin del experimento en el ensayo 32. La gráfica inferior de la Figura 8(a) refleja este evidente cambio sistemático y gradual en la orientación del robot desde el giro a -90° hasta el giro a $+90^\circ$.

El criterio en el laberinto T y en el radial ocurren alrededor del mismo ensayo, así que, al igual que en el caso de los robots "sanos", el cambio abrupto del giro hacia el brazo izquierdo de la T al giro a la derecha a partir del ensayo 13 es el reflejo del momento en que la orientación del robot en el laberinto radial cruza la línea media.

El desempeño promedio del robot "lesionado" muestra una similitud cualitativa con los resultados obtenidos por O'Keefe al experimentar con seis ratas lesionadas en el laberinto radial. La gráfica inferior de la Figura V.8(b) extraída y adaptada de (O'Keefe, 1983), presenta 32 ensayos de prueba. El criterio en este caso ocurre entre el ensayo 16 y el 20, donde la orientación promedio de las ratas cruza la línea media pasando del cuadrante de rotaciones negativas al cuadrante de rotaciones positivas. Como en el caso de los robots, las ratas eligen el nuevo giro con recompensa consistentemente a partir del ensayo 24, y la gráfica ilustra un cambio sistemático y gradual del giro originalmente recompensado (-90°) al nuevo giro que ofrece la recompensa ($+90^\circ$).

V.1.3.3. El Laberinto de Múltiples Ts con Ratas Sanas

Después de implementar la tarea de discriminación inversa de O'Keefe utilizando un laberinto tipo T y uno radial de 8 brazos, extendimos el experimento considerando un laberinto más complejo, donde cualquier ruta que explore un robot "sano" a partir de una localidad específica incluye dos puntos de decisión. La forma diseñada para dicho laberinto involucra dos Ts horizontales ancladas en los brazos de una T vertical como lo muestra la Figura V.9(a).

Aún cuando el trabajo propuesto por Guazzelli y colaboradores (Guazzelli et al., 1998) también fue validado mediante la reproducción de los resultados obtenidos por O'Keefe en los laberintos T y radial, la tarea inversa en su caso fue implementada únicamente dentro de un ambiente simulado no extensible a laberintos más complejos.

En cada ensayo del experimento en el laberinto de múltiples Ts, el robot navega de la base de la T vertical al fondo de cualquiera de sus brazos (0° ó 180°), y de ahí al fondo de cualquiera de los brazos de la T horizontal correspondiente (90° ó 270°). Después, el robot regresa a la localidad de partida de manera autónoma. Durante el entrenamiento del robot, la meta se encuentra localizada al fondo del brazo derecho (90°) de la T horizontal izquierda (180°), mientras que durante las pruebas, la localización de la meta cambia al fondo del brazo derecho (270°) de la T horizontal derecha (0°).

Al inicio del entrenamiento, la exploración del laberinto por parte del robot está determinada por curiosidad y ruido. El nivel de curiosidad por rotar -90° ó $+90^\circ$ en cualquier localidad de decisión decrece después de haber visitado una vez el corredor correspondiente, prevaleciendo entonces el ruido.

El mapa construido durante el proceso de exploración se muestra en la Figura V.9(b). Al término de cualquier ensayo, el proceso de refuerzo hacia atrás tiene lugar sobre los nodos del mapa que pertenecen a la ruta seguida por el robot. Como se describió en la Sección III.8, este proceso consiste en actualizar los rastros de elegibilidad de las unidades Actor asociadas a las direcciones de los arcos entre nodos. En caso de que el robot haya alcanzado la meta al final del ensayo, la ruta es reforzada positivamente (aprendizaje de ruta); de otro modo, la trayectoria es reforzada negativamente (desaprendizaje de ruta).

Eventualmente, el robot alcanza el criterio una vez que la expectativa por encontrar una recompensa al orientarse a 180° en la primera localidad de decisión y a 90° en la segunda se vuelve mayor que el ruido. Cuando este evento ocurre, la fase de entrenamiento concluye. La duración promedio del entrenamiento obtenida a partir de seis ejecuciones del modelo fue de 13 ensayos, y considerando que un ensayo tardó 2.5 minutos en promedio, el robot completó la fase de entrenamiento en un poco más de 30 minutos.

Durante la tarea inversa ocurre el proceso de desaprendizaje. Cuando la expectativa de recompensa por la ruta aprendida durante el entrenamiento llega a ser menor que el ruido, el robot comienza a ejecutar acciones aleatorias en las localidades de decisión del laberinto hasta que alcanza el criterio una vez que la expectativa de recompensa por orientarse a 0° en la primera localidad de decisión y a 270° en la segunda, se vuelve mayor al ruido. Cuando esto ocurre, el robot ha aprendido la nueva ruta. El desempeño promedio del robot durante la tarea inversa derivado de seis ejecuciones del modelo se presenta en la Figura V.10, donde la conducta del robot está expresada en términos del porcentaje de decisiones correctas en ambas localidades de decisión del laberinto.

Como resultado del entrenamiento, el 100% de las elecciones del robot fueron correctas (control). La gráfica de la Figura V.10 muestra 32 ensayos de prueba. El robot tomó 12 ensayos en desaprender la ruta previamente correcta hacia la meta (criterio). Como se esperaba, es posible apreciar un cambio abrupto de la ruta incorrecta a la correcta de la misma manera que sucedió en el laberinto T simple. Así, el porcentaje de elecciones correctas del robot pasó de 32% en el ensayo 12 a 90% en el ensayo 16 y de ahí al 100%.

Como en el caso de la tarea inversa en los laberintos tipo T y radial, el comportamiento del robot en el laberinto de múltiples Ts depende de la representación espacial construida en base a la información kinestésica, dado que el robot no dispuso de landmarks externos durante la tarea.

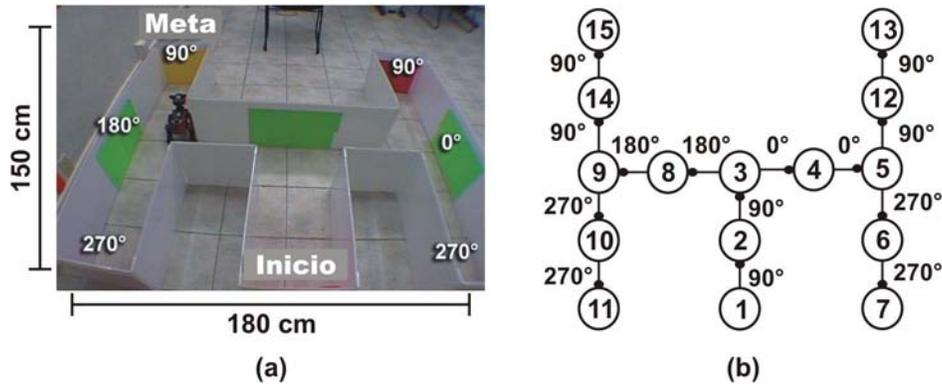


Figura V.9. (a) El laberinto extendido utilizado en la tarea de discriminación inversa. La meta se presenta en la localidad que corresponde durante el entrenamiento del robot. (b) El mapa construido por el robot durante el proceso de exploración. Los nodos se muestran numerados en orden de creación, y los arcos entre nodos están asociados a la dirección aloentrónica del robot cuando se desplazó de un nodo al siguiente.

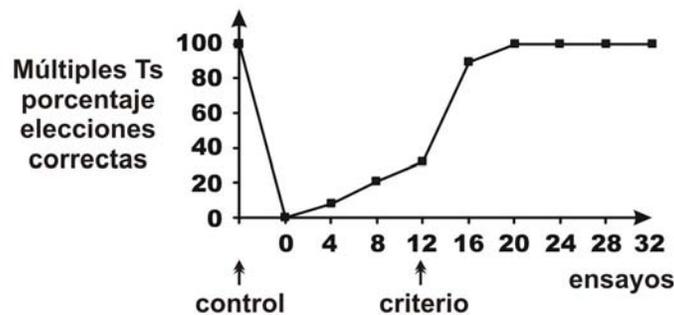


Figura V.10. El desempeño promedio de seis robots expresado en términos del porcentaje de decisiones correctas tomadas durante la tarea inversa en el laberinto de múltiples Ts. La gráfica fue obtenida promediando las gráficas de los robots individuales en periodos de cuatro ensayos. Es posible apreciar un cambio abrupto de la ruta incorrecta a la correcta.

V.2. NAVEGACIÓN ORIENTADA A METAS

De acuerdo con la teoría del mapa cognitivo (O'Keefe y Nadel, 1978), la utilidad que éste provee durante el aprendizaje de laberintos incluye localizar la posición del animal dentro del ambiente y señalar la dirección de la meta desde tal posición. Esta información habilita al animal para manifestar una conducta de navegación orientada a metas en la que es factible alcanzar la localidad meta desde cualquier otra localidad del laberinto aprendido. Una de las tareas espaciales clásicas que evidencian dicho comportamiento es el laberinto de agua, introducido por Morris (1981), donde se observó que las ratas sanas pueden nadar hacia una plataforma sumergida independientemente de la localidad a partir de la cual sean liberadas.

Esta sección describe la tarea espacial en el laberinto de agua como antecedente para presentar los resultados obtenidos con robots al experimentar con una tarea de protocolo similar en un laberinto en tierra rodeado por landmarks.

V.2.1. El Laberinto de Agua de Morris

En la tarea diseñada por Morris (1981, 1983), ratas sanas y ratas con lesiones en el hipocampo fueron colocadas de forma independiente en un tanque circular con diámetro de 1.32 m lleno de un líquido opaco resultante de mezclar agua con leche. Las ratas debían nadar en este tanque hasta localizar una plataforma hecha de acrílico transparente colocada en una posición fija en la que pudieran subir para escapar del agua fría ($26^{\circ} \pm 1^{\circ}\text{C}$).

Los animales fueron probados en dos situaciones: (i) con la plataforma visible, estando elevada 1 cm sobre la superficie del agua, por lo que las ratas sólo tenían que aprender que la plataforma les permitiría escapar del agua fría, y (ii) con la plataforma escondida justo debajo de la superficie del agua, donde las ratas igualmente podrían subir para escapar, pero al no estar visible, sólo escaparían rápidamente si aprendían la localización de la plataforma. En esta segunda situación, señales visuales distantes fueron colocadas alrededor del laberinto.

Después del entrenamiento, todos los animales fueron capaces de nadar directamente hacia la plataforma visible. No obstante, en el caso de la plataforma sumergida, sólo las ratas sanas la encontraron desde cualquier localidad de inicio en la periferia del tanque. La Figura V.11 ilustra las trayectorias seguidas por ratas sanas en la condición de la plataforma escondida.

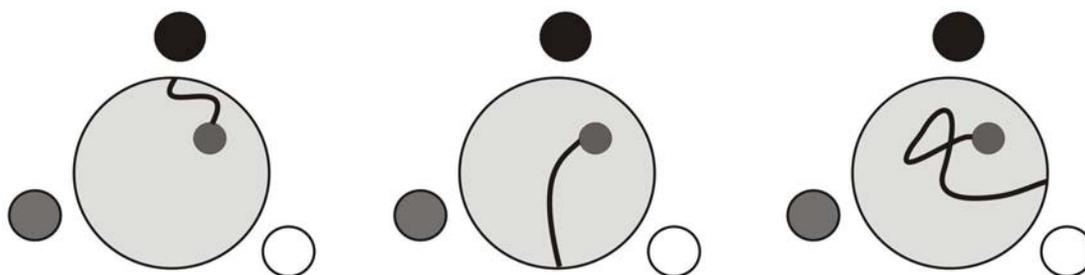


Figura V.11. Trayectorias seguidas por las ratas sanas para encontrar la plataforma sumergida en el tanque de agua de Morris. Tres landmarks rodean el tanque, y la plataforma permanece en el cuadrante noreste del tanque durante todo el experimento. Después del entrenamiento, los animales alcanzan la localidad de la plataforma independientemente de la localidad de liberación: el lado norte del tanque, el lado sur y el lado este. Diagramas adaptados de (Roberts, W. A., 1998).

Si después del entrenamiento la plataforma era colocada en otro lugar o removida del tanque, las ratas sanas buscaban la plataforma persistentemente en la ubicación previa. En otro caso, si la plataforma era colocada en un lugar aleatorio distinto de un ensayo a otro de entrenamiento, las ratas aprendían a escapar del agua, pero no a navegar directamente hacia la plataforma. La navegación hacia la posición correcta en el espacio dependió, entonces, de la permanencia de la meta en una localización fija y relativa a señales visuales distantes.

En síntesis, los animales con hipocampo sano son capaces de resolver la tarea empleando una estrategia de aprendizaje de lugares, la cual les permite relacionar la posición de la meta con la de landmarks externos, reconocer la localidad de la meta dentro del mapa cognitivo, y navegar hacia ella. En cambio, los animales con hipocampo lesionado sólo están habilitados para utilizar una hipótesis de orientación, la cual manifiestan aprendiendo el procedimiento que conduce a la meta o navegando directamente hacia una meta visible.

V.2.2. Resultados Experimentales

En la tarea espacial de discriminación inversa en un laberinto de múltiples Ts con ratas sanas que fue descrita en la Sección V.1.3.3, el objetivo primario consistió en probar la habilidad del robot para aprender cómo navegar hacia la meta desde una localidad fija utilizando solamente información kinestésica, y para desaprender la localidad de la meta cuando ésta es cambiada a otra parte del laberinto. En el experimento que presenta esta sección, tres cilindros de colores representando landmarks fueron dispuestos alrededor del laberinto de múltiples Ts como lo muestra la Figura V.12(a), con la finalidad de probar: (i) el proceso de representación y reconocimiento de lugares llevado a cabo por el robot empleando no sólo información kinestésica sino también visual mientras explora el laberinto, y (ii) la navegación orientada a metas que permite al robot alcanzar la meta desde distintas localidades de inicio.

El protocolo del experimento efectuado en el laberinto rodeado por landmarks está inspirado en la tarea espacial de Morris (1981) y adaptado para un laberinto en tierra con corredores. De hecho, una versión basada en tierra de esta tarea fue previamente implementada por Hollup y colaboradores (2001(2)) con el objetivo de determinar que, a pesar de las diferencias relativas a la información kinestésica de entrada, los campos de lugar hipocampales en un laberinto basado en tierra con un corredor circular colocado al centro son controlados por los mismos factores que intervienen en un laberinto de agua abierto y en el laberinto de agua incluyendo un corredor en su centro.

La fase de entrenamiento procede como la tarea de discriminación discutida en la Sección V.1.3.3, esto es, en cada ensayo, el robot parte de una localidad fija previamente designada, explora el laberinto hasta encontrar la meta o el final de un corredor, y regresa al punto de inicio. Durante la exploración, el robot construye un mapa similar al que se muestra en la Figura V.9(b).

Antes de implementar la tarea, decidimos ajustar algunos de los parámetros del modelo relacionados con el proceso de refuerzo hacia atrás con el afán de reducir la duración total del entrenamiento del robot, aún cuando sabemos que las ratas requieren bastantes ensayos para aprender la localidad de la meta en cualquier laberinto. Como resultado, el robot requiere experimentar en promedio únicamente cinco ensayos de entrenamiento encontrando la meta para aprender esta localidad. La ejecución del modelo en seis ocasiones permitió determinar una duración promedio de la fase de entrenamiento de 9 ensayos equivalentes a 23 minutos aproximadamente.

Después del entrenamiento, la localidad meta permaneció constante, pero el robot fue colocado en diferentes posiciones de partida durante los ensayos de prueba. En la mejor de las seis ejecuciones del modelo, el robot logró seguir rutas directas hacia la meta como lo ilustra la Figura V.12(b), donde se indican las localidades de partida del robot, D1, D2 y D3, para tres ensayos. La Figura V.12(c) muestra el mapa del ambiente actualizado por el robot durante las pruebas. Conviene destacar que los seis robots (esto es, las seis ejecuciones del modelo) lograron encontrar la meta desde todas las posiciones de inicio probadas, aún cuando algunos tomaron caminos más largos desde las localidades D2 y D3. En promedio, las duraciones de los ensayos que iniciaron en D1, D2 y D3 fueron 50, 120 y 140 segundos, respectivamente.

Para alcanzar el objetivo desde cualquier localidad, el robot ejecuta los procesos de reconocimiento de lugares y de explotación del mapa cognitivo. Considérese, por ejemplo, el ensayo más simple que inicia en la localidad D1 (ver Figura V.12(b-c)). El robot inicia la prueba estando orientado a 90° (norte). Dado que no existen landmarks visibles desde ahí, las células de lugar del modelo responden por medio de información kinestésica únicamente. La actividad global que actualmente registran las células es buscada dentro de los nodos del mapa y encontrada dentro de la unidad Actor 270° conectada al nodo 11. Aunque la dirección actual del robot no es 270° (sur), ambos patrones de actividad son similares dado que ninguno de ellos codifica información visual. Por lo tanto, el nodo 11 se convierte en el nodo activo dentro del mapa, indicando de esta forma que el robot reconoce la localidad D1 como una previamente visitada. Después, el robot se desplaza hacia delante y percibe una parte del landmark 3 (L3). Esta vez, el patrón de actividad actual derivado de la colección de células de lugar no resulta similar a ninguno de los patrones almacenados en el

mapa. Esto es así porque el robot visitó previamente esa localidad estando orientado solamente a 270° y no a 90° , por lo que la información visual percibida en ambas direcciones es diferente. En consecuencia, el patrón de actividad actual es almacenado en una nueva unidad Actor 90° , la cual es asociada con el nuevo nodo número 16. Cuando el robot alcanza la localidad de decisión del corredor, la actividad conjunta registrada actualmente por las células de lugar es encontrada en la unidad Actor 90° del nodo 9, así que este nodo es activado en el mapa. En esta localidad, la expectativa de futura recompensa en dirección 90° fue incrementada durante el entrenamiento de manera importante; por lo tanto, el robot explota dicha información cuando decide avanzar hacia delante en vez de girar a la derecha. A partir de esta localidad, el robot se desplaza hacia delante mientras reconoce los lugares representados en el mapa por el nodo 14 hasta que llega a la meta, cuya localidad es representada por el nodo 15.

Los procesos de reconocimiento de lugares y explotación del mapa cognitivo efectuados por el robot para alcanzar la meta durante los ensayos que iniciaron en las localidades D2 y D3 son similares a los recién descritos para el caso que inicia en D1. En estas dos situaciones, el robot agrega al mapa los nodos 17, 18 y 19, y emplea los valores de expectativas de recompensa para decidir avanzar hacia delante en la localidad representada por el nodo 3 y girar a la derecha en la localidad representada por el nodo 9 cuando está orientado a 180° (oeste).

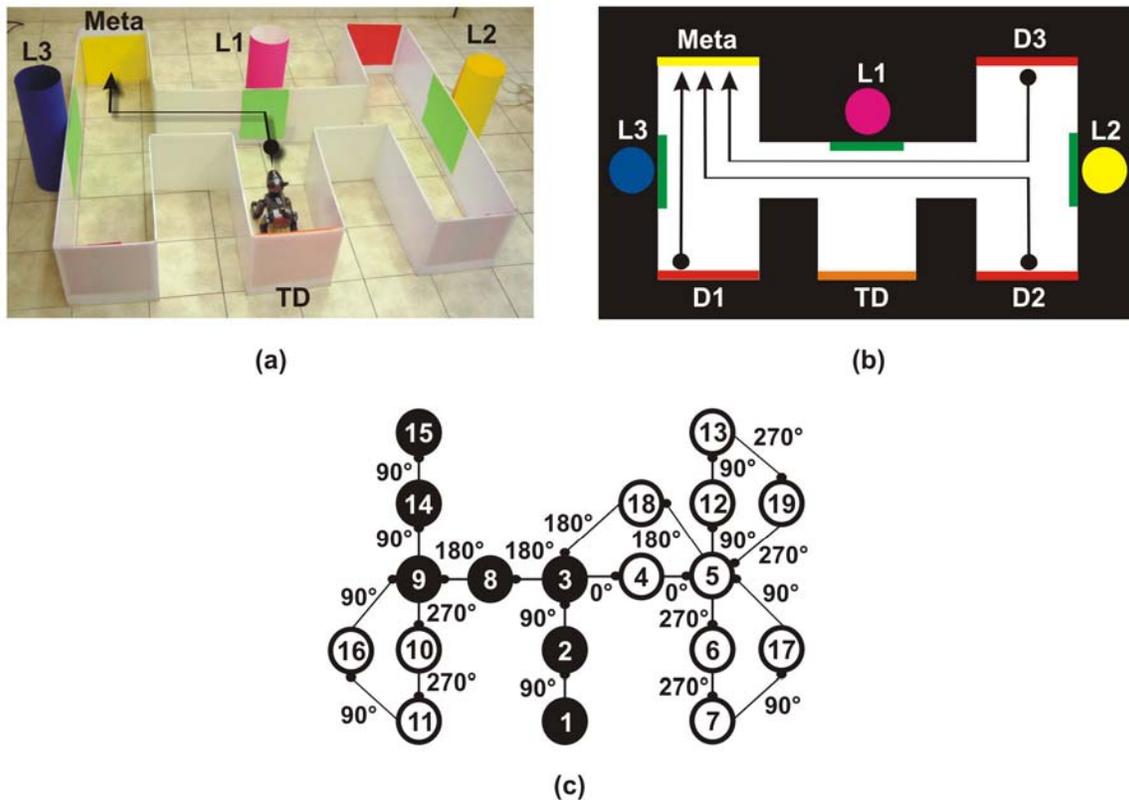


Figura V.12. (a) El laberinto de múltiples Ts rodeado por tres landmarks identificadas como L1, L2 y L3. El robot Sony AIBO se muestra localizado en la posición de partida durante el entrenamiento (TD), y el camino hacia la meta se presenta marcado con una flecha. (b) Rutas seguidas por el robot en tres ensayos de prueba que inician en las localidades identificadas como D1, D2 y D3. (c) El mapa cognitivo actualizado por el robot durante la fase de prueba. Los nodos están numerados en orden de creación. Los nodos negros pertenecen a la ruta aprendida durante el entrenamiento.

Una comparación apropiada entre los resultados obtenidos con robots y los reportados por Morris al experimentar con ratas sanas resulta difícil dadas las importantes variaciones implementadas al experimento original. No obstante, como lo hizo Morris con las ratas sanas, fue posible confirmar en el caso de los robots que para encontrar una meta “escondida” en el laberinto independientemente de la localidad de inicio, el robot

requiere explotar su mapa cognitivo reconociendo las expectativas de recompensa que lo hacen navegar hacia la localidad objetivo exitosamente.

V.3. ADAPTACIÓN DEL MAPA COGNITIVO ANTE CAMBIOS FÍSICOS

La teoría postulada por O'Keefe y Nadel (1978) en relación al mapa cognitivo asume que éste es adaptado por los animales como resultado del proceso de desaprendizaje de localidades meta y en función del nuevo aprendizaje. El modelo propuesto de cognición espacial y navegación en ratas considera extensiva dicha adaptación del mapa ante situaciones en las que el ambiente cambia su configuración física durante la exploración por parte del animal.

Mientras el robot explora el ambiente, cambios físicos en su configuración implican cambios en la información visual que el robot percibe. La entrada de nueva información visual desde localidades diversas en el ambiente puede provocar la eliminación de los nodos que actualmente las representan dentro el mapa cognitivo y la creación de nuevos nodos que efectivamente reflejen la percepción actual del robot, o bien, la integración de nodos existentes. En cualquier caso, el ajuste del mapa puede involucrar que la información relativa a las expectativas del robot por encontrar recompensas en localidades específicas sea reasignada a nuevos nodos.

Concretamente, el modelo distingue entre dos situaciones de adaptación del mapa cognitivo durante el proceso de exploración del ambiente:

1. Integración de nodos. Cuando el robot gira estando en una misma localidad y se activa en el mapa un nodo X diferente al nodo activo, el nodo X y el activo deben ser integrados en un solo nodo, pues ambos representan en el mapa la misma localidad del ambiente.

La integración de dos nodos implica la creación de un nuevo nodo que incorpora todos los arcos que salen del nodo X y del activo, así como todas las unidades Actor conectadas a cada uno, las cuales están asociadas con direcciones diferentes. No es posible encontrar en ambos nodos dos unidades Actor asociadas con la misma dirección porque en cuanto se detecta esta situación los nodos se mezclan. De igual forma, todos los arcos que apuntan al nodo X y al activo son actualizados para apuntar al nuevo nodo, el cual ahora es el activo. Finalmente, los nodos integrados son eliminados del mapa. La Figura V.13(a) ilustra la situación descrita utilizando una muestra del mapa de un laberinto en forma de cruz.

2. Creación / eliminación de nodos. Este proceso involucra dos casos:
 - (i) Cuando se crea un nuevo nodo y el nodo activo se conecta a éste a través de un arco asociado con en cierta dirección d , y ya existe un arco que sale del nodo activo en la misma dirección d , este arco se elimina, pues no es factible llegar a dos lugares diferentes desde la localidad representada por el nodo activo siguiendo la misma dirección. La Figura V.13(b) muestra este caso gráficamente empleando únicamente los 3 nodos implicados en el procedimiento.
 - (ii) Cuando se crea un nuevo nodo X en el instante de tiempo $t-1$, se activa un nodo existente Y en el tiempo t , se conecta el nodo X al nodo activo Y a través de un arco asociado con cierta dirección d , y ya existe un nodo Z con un arco que apunta al nodo activo Y en la misma dirección d , este arco se elimina, pues no es posible llegar a la localidad representada por el nodo activo Y desde dos lugares diferentes (nodos X y Z) siguiendo la misma dirección. La unidad Actor asociada con la dirección d del nodo Z es asignada ahora al nodo X incluyendo el valor de la expectativa de recompensa existente. La Figura V.13(c) ilustra esta situación con tres nodos.

En cualquiera de estos casos (i) y (ii), los nodos que ya no tengan arcos salientes ni entrantes como resultado del ajuste del mapa, son eliminados.

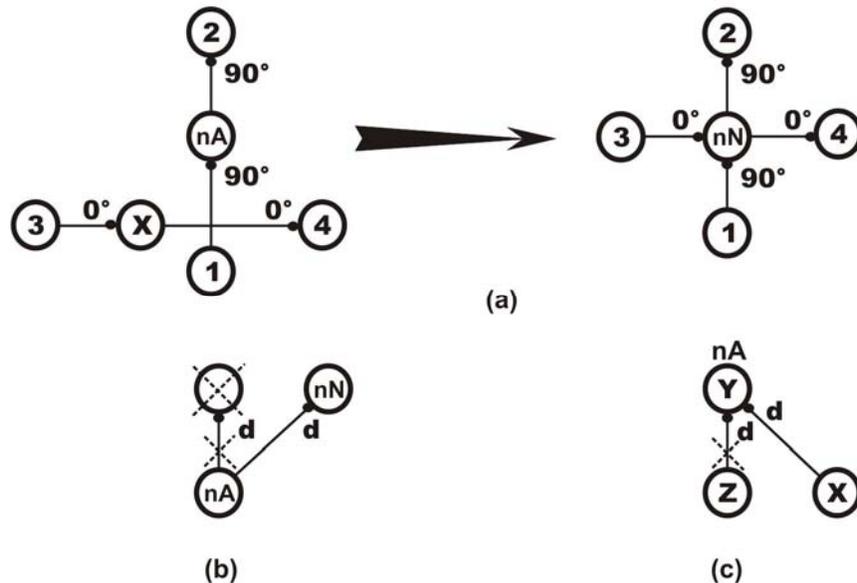


Figura V.13. Esquematación de las situaciones de adaptación del mapa cognitivo durante el proceso de exploración del ambiente. Glosario: nA – nodo activo; nN – nuevo nodo. (a) La integración de los nodos nA y X en un solo nodo nN . (b) La creación de nN , la conexión de nA a nN en dirección d , y la eliminación del arco existente en dirección d de nA a otro nodo. (c) La conexión del nodo X al nodo activo Y en dirección d , la eliminación del arco existente en dirección d del nodo Z al Y , y la asignación de la unidad Actor d del nodo Z al nodo X .

Las situaciones descritas fueron probadas con el robot en el laberinto de múltiples Ts rodeado por landmarks alterando la disposición física de estas señales visuales después de haber entrenado al robot para encontrar la meta. En particular, se experimentó separadamente con dos escenarios: el intercambio y la eliminación de landmarks en el ambiente. Los próximos apartados discuten los resultados obtenidos en ambos casos. Una descripción sintética de los mismos fue reportada en (Barrera y Weitzenfeld, 2007).

V.3.1. Intercambio de Landmarks

Después de entrenar al robot en el laberinto de múltiples Ts rodeado por landmarks que muestra la Figura V.12(a), observando el mismo protocolo descrito en la Sección V.2.2, los landmarks L2 y L3 fueron intercambiados. Posteriormente, el robot fue probado partiendo de la localidad de entrenamiento y de localidades diferentes a ésta. La Figura V.14(a) ilustra dos de las localidades de salida experimentadas por el robot (TD y D2), así como ejemplos de las trayectorias que siguió. El mapa resultante del experimento es mostrado en la Figura V.14(b), donde los nodos negros fueron creados durante el entrenamiento desde la localidad TD, y los nodos blancos fueron agregados durante las pruebas desde localidades diversas.

Durante las pruebas que inician en las localidades identificadas como D2 y TD en la Figura V.14(a), el robot reconoce los lugares desde donde no fue posible percibir los landmarks L2 y L3 durante el entrenamiento, tales como las localidades representadas en el mapa por los nodos 1, 2, 6, 7, 10 y 11. En cambio, no reconoce los lugares desde donde L2 y L3 fueron visibles en el entrenamiento, por lo que agrega nuevos nodos en el mapa para representarlos. Éste es el caso de los lugares comprendidos en el corredor horizontal del laberinto, y para representarlos en el mapa, el robot crea los nodos 17, 18, 19, 20, 21 y 22. La creación de nuevos nodos y arcos en el mapa implica la eliminación de algunos arcos existentes como lo muestra la Figura V.14(b). Dado que los nodos 19 y 22, creados para reemplazar los nodos de decisión 3 y 9,

no están asociados con expectativas de recompensa en alguna dirección, el robot logra alcanzar la localidad meta ejecutando únicamente acciones aleatorias.

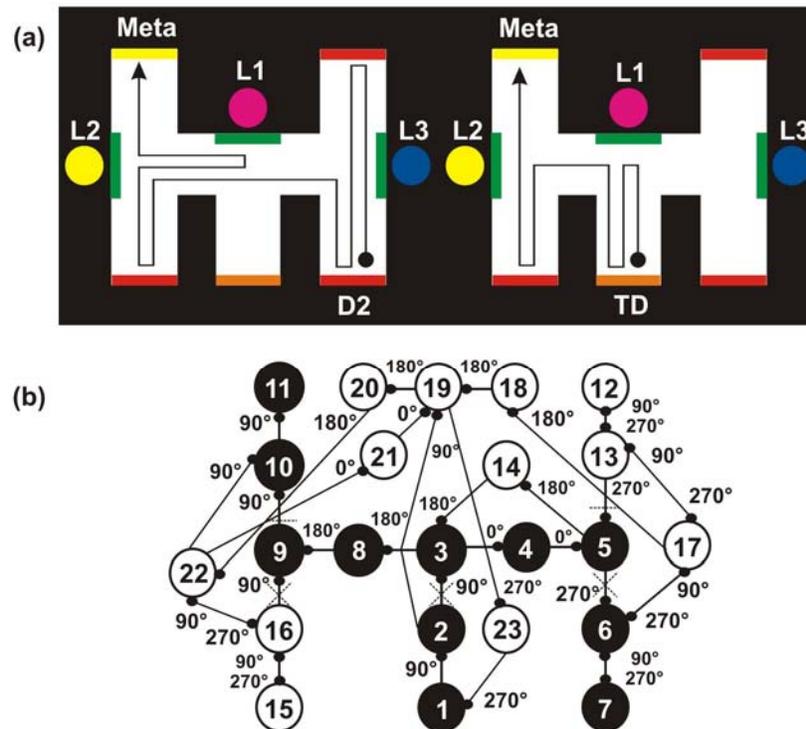


Figura V.14. (a) Trayectorias seguidas por el robot en las pruebas iniciadas en las localidades D2 y TD, donde los landmarks L2 y L3 fueron intercambiados. (b) El mapa construido por el robot durante el experimento. Los nodos son identificados por números y los arcos están asociados con la dirección del robot cuando éste se movió de un nodo al siguiente. Los nodos negros fueron creados durante el entrenamiento, y los blancos fueron agregados en las pruebas. Los arcos marcados con "X" fueron eliminados durante las pruebas.

V.3.2. Eliminación de Landmarks

El robot fue probado nuevamente partiendo de la localidad de entrenamiento y de localidades diferentes a ésta después de retirar del laberinto uno de los landmarks. Los resultados obtenidos son descritos a continuación.

V.3.2.1. Removiendo L1

La Figura V.15(a) muestra ejemplos de trayectorias seguidas por el robot durante las pruebas desde D2 y TD, donde L1 fue retirado del laberinto. El mapa resultante del experimento es mostrado en la Figura V.15(b), donde los nodos negros fueron creados durante el entrenamiento.

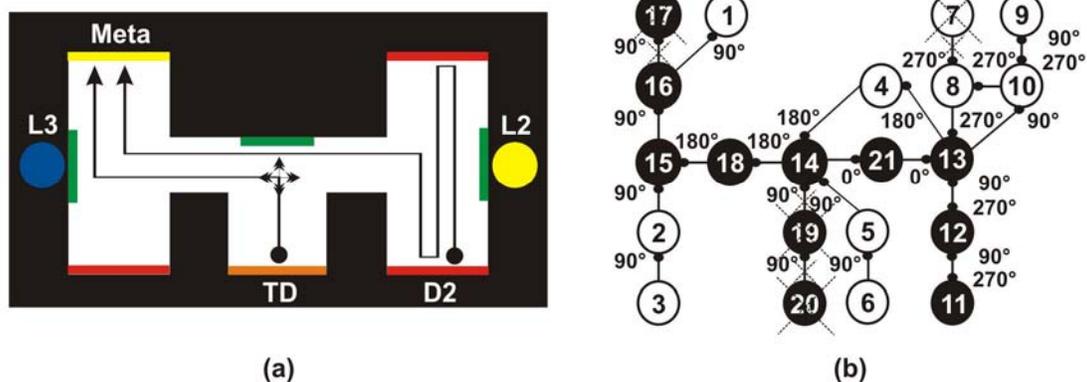


Figura V.15. (a) Rutas seguidas por el robot durante las pruebas iniciadas en las localidades TD y D2, donde el landmark L1 fue retirado (b) El mapa construido por el robot durante el experimento. Los nodos negros fueron creados durante el entrenamiento, y los blancos durante las pruebas. Los nodos y los arcos marcados con "X" fueron eliminados durante las pruebas.

Prueba desde D2. El robot establece conexiones entre los nodos 11-12 y 12-13 en dirección 90° , reconociendo estos lugares en virtud de que L1 no fue visible desde ahí durante el entrenamiento. El nodo 8 fue previamente visitado en dirección 270° , así que el nuevo nodo 10 es creado y vinculado con el nodo 13 en la dirección actual 90° . Cuando el robot llega al final del corredor y se orienta a 270° , el patrón de activación actual de las células de lugar del modelo no es similar al que se encuentra almacenado en la unidad Actor que corresponde a la dirección 270° del nodo 7 debido a la ausencia de L1. Por lo tanto, el nodo 9 es creado, las conexiones 9-10 y 10-8 son establecidas en dirección 270° , y la liga 7-8 es eliminada dado que no es posible alcanzar el nodo 8 desde dos lugares distintos (nodos 7 y 10) en la misma dirección. El robot sigue la ruta correcta a la meta una vez que alcanza la localidad representada por el nodo 14 en dirección 180° . El nodo 14 es reconocido porque el patrón visual almacenado en el Actor 180° incluye información de L3 además de L1. Aunque el robot llega a la meta, no reconoce el nodo 17 por la ausencia de L1, así que agrega el nodo 1. La conexión 16-17 en dirección 90° es eliminada, así como el nodo 17, y la conexión 16-1 es creada.

Prueba desde TD. El robot no reconoce los lugares representados por los nodos 20 y 19, pues sus unidades Actor en dirección 90° almacenan patrones visuales con información referente a L1 solamente. Por tanto, los nodos 6 y 5 son creados, agregados al mapa, y vinculados en dirección 90° . Cuando el robot alcanza la localidad representada por el nodo 14, no reconoce el lugar inmediatamente; sin embargo, lo hace cuando elige aleatoriamente girar a la derecha tres veces para orientarse finalmente hacia la dirección 180° . Entonces, la conexión 5-14 es creada en dirección 90° ; el Actor 90° del nodo 19 es asignado al nodo 5, y el Actor 90° del nodo 20 es asignado al nodo 6. Las conexiones entre los nodos 19-14 y 20-19 son eliminadas, al igual que los nodos 19 y 20. A partir del nodo 14, el robot sigue un camino directo a la meta.

V.3.2.2. Removiendo L2

La Figura V.16(a) muestra ejemplos de trayectorias seguidas por el robot durante las pruebas desde D2 y TD, donde L2 fue retirado del laberinto. El mapa resultante del experimento es mostrado en la Figura V.16(b), donde los nodos negros fueron creados durante el entrenamiento.

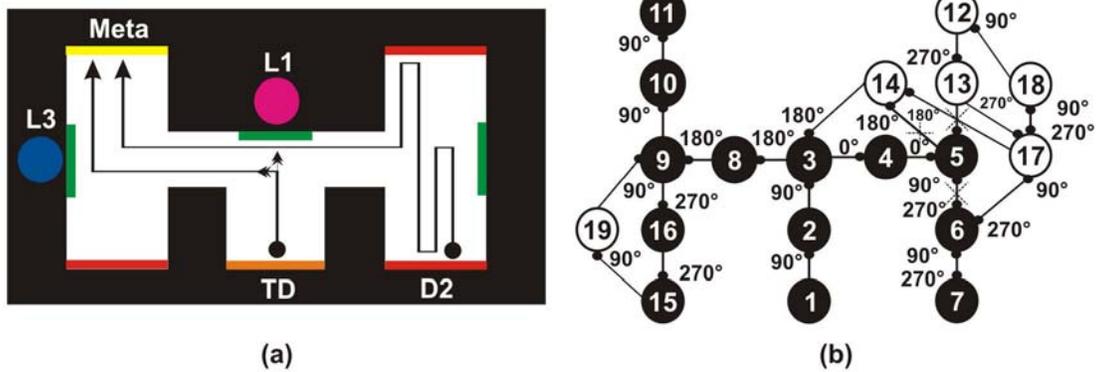


Figura V.16. (a) Rutas seguidas por el robot durante las pruebas iniciadas en las localidades TD y D2, donde el landmark L2 fue removido. (b) El mapa construido por el robot durante el experimento. Los nodos negros fueron creados durante el entrenamiento, y los blancos en las pruebas. Los arcos marcados con "X" fueron eliminados durante las pruebas.

Prueba desde D2. El robot no reconoce la localidad representada por el nodo 5. La unidad Actor 90° de este nodo almacena información principalmente relacionada con L2, y, aunque también incluye información sobre L3, el patrón de activación actual de las células de lugar no es similar al almacenado en dicha unidad Actor. De aquí que el nodo 17 sea creado y conectado al nodo 6 en dirección 90° , y la conexión 6-5 en esta dirección sea eliminada. Posteriormente, el robot elige girar a la izquierda dos veces orientándose a 270° , y se define entonces la conexión 17-6 en dirección 270° . El Actor 270° del nodo 5 es asignado al nodo 17, y la conexión 6-5 en dirección 270° es removida. Estando orientado nuevamente a 90° , el robot alcanza el nodo 17, y como no reconoce el nodo 13, el nodo 18 es agregado al mapa y conectado al 17 en dirección 90° . El robot reconoce el nodo 12 pues L2 no fue visible desde ahí durante el entrenamiento, y la conexión 18-12 es definida. Cuando el robot regresa, la conexión 13-17 se agrega en dirección 270° , y la conexión 13-5 es eliminada. Finalmente, cuando el robot gira a la derecha orientándose a 180° , la liga 17-14 es creada. El Actor 180° del nodo 5 es asignada al nodo 17, y la conexión 5-14 es eliminada. Después, el robot reconoce el nodo 3 y sigue el resto de la ruta correcta hacia la meta.

Prueba desde TD. Estando en dirección 90° , el robot reconoce los nodos 1 y 2 inmediatamente, pero no así el nodo 3, el cual es reconocido sólo cuando decide aleatoriamente girar a la izquierda orientándose a 180° . En este punto, la ausencia de L2 no es relevante dentro del patrón de información visual almacenado en el Actor 180° del nodo 3. A partir de esta localidad, el robot llega a la meta directamente.

V.3.2.3. Removiendo L3

La Figura V.17(a) muestra ejemplos de trayectorias seguidas por el robot durante las pruebas desde D2 y TD, donde L3 fue retirado del laberinto. El mapa resultante del experimento es mostrado en la Figura V.17(b), donde los nodos negros fueron creados durante el entrenamiento.

V.4. RECONOCIMIENTO DE LUGARES EN MÚLTIPLES DIRECCIONES

Cuando termina el entrenamiento del robot para aprender una tarea espacial, el mapa construido por éste incluye nodos que representan las localidades exploradas, y unidades Actor asociadas a cada nodo que corresponden a las diversas direcciones en que el robot exploró la localidad. Sin embargo, durante el entrenamiento, el robot no necesariamente visita todas las localidades del laberinto ni en todas las posibles direcciones, en parte porque aprende relativamente rápido y también porque el modelo restringe las rotaciones de 180° ; así que durante las pruebas, es factible que el robot agregue nodos para representar localidades exploradas por primera vez y/o localidades previamente exploradas, pero en direcciones experimentadas por primera vez. Como se describió en la Sección V.3, estos múltiples nodos que representan un mismo lugar sólo pueden llegar a integrarse cuando el robot gira estando en tal lugar y se activa un nodo diferente al actual; pero si esto no ocurre en todos los casos de multiplicidad, el mapa final no revelará la organización del espacio con precisión, ni reflejará la capacidad del robot para reconocer una misma localidad estando orientado hacia direcciones distintas.

Por otro lado, de acuerdo con la conducta observada de las ratas, éstas suelen habituarse naturalmente a un nuevo ambiente, ya sea antes o durante la fase de entrenamiento. En efecto, avanzan, giran y regresan explorando el laberinto en su totalidad, y cuando ya lo conocen y logran aprender la tarea espacial, se desplazan de forma rápida y siguen rutas directas hacia la meta en la mayoría de los casos.

Con la finalidad de mejorar durante las pruebas la adaptación del mapa cognitivo resultante del entrenamiento, y por ende el reconocimiento de lugares por parte del robot, se introdujo una fase de habituación al ambiente antes de iniciar el proceso de aprendizaje de la tarea espacial. Específicamente, este aspecto fue probado al experimentar con el robot en el laberinto cíclico mostrado en la Figura V.18(a), donde cuatro cilindros de colores fueron utilizados para representar landmarks externos. En este laberinto, como en los experimentos presentados en secciones anteriores, el robot debe aprender la localidad meta desde un punto de partida fijo de tal forma que después sea capaz de navegar hacia ella desde posiciones de partida distintas. La misma tarea espacial fue implementada con un grupo de ratas sanas en el laberinto mostrado en la Figura V.18(b), con el afán de establecer un análisis comparativo entre los resultados obtenidos con las ratas y con el robot.

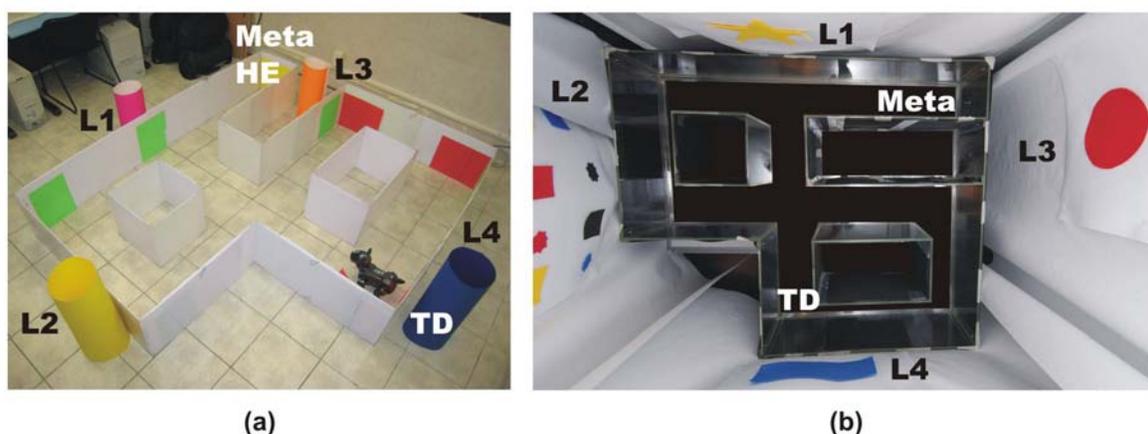


Figura V.18. (a) El laberinto cíclico empleado en el experimento con el robot Sony AIBO, el cual se presenta colocado en el punto de partida de entrenamiento TD. Los landmarks identificados por L1, L2, L3 y L4 se muestran dispuestos fuera del laberinto, y la meta está situada en la localidad HE. (b) El laberinto cíclico equivalente empleado en el experimento con ratas.

En los siguientes apartados se describe la conducta promedio observada en el grupo de ratas resolviendo la tarea espacial en el laberinto cíclico, así como el desempeño del robot al aprender la misma tarea. Adicionalmente, se presentan los resultados derivados del análisis comparativo entre el comportamiento de las ratas y del robot.

V.4.1. El Laberinto Cíclico: Resultados Experimentales con Ratas

Se utilizó un grupo de tres ratas sanas para llevar a cabo el experimento en el laberinto mostrado en la Figura V.18(b). Con la intención de motivar a los animales para que pudieran aprender la tarea espacial, éstos fueron absolutamente privados de agua durante la semana previa al experimento. La punta de un bebedero fue dispuesta en la localidad meta indicada en la figura para proveer agua dulce a los animales cuando alcanzaran dicho lugar al terminar un ensayo del experimento. La solución utilizada como reforzador consistió de 5 gramos de sacarosa disueltos en 50 mililitros de agua, y 400 microlitros de la solución fueron suministrados en el bebedero antes de comenzar cualquier ensayo.

Los siguientes apartados describen la conducta de las ratas durante las fases de entrenamiento y prueba del experimento.

V.4.1.1. Fase de Entrenamiento

Al inicio de cualquier ensayo de entrenamiento, la rata es extraída de su caja de acrílico y cubierta con una toalla de tal manera que no pueda percibir información ajena a la tarea espacial mientras es transportada al laberinto de la Figura V.18(b). Después, la rata es colocada en la localidad TD estando orientada hacia el norte (90°), y el ensayo consiste en explorar libremente el laberinto hasta alcanzar la localidad meta y consumir el agua dulce. Al terminar el ensayo, la rata es extraída del laberinto, cubierta nuevamente con una toalla y depositada en su caja.

La duración del entrenamiento dependió del momento en que los animales alcanzaran el criterio de aprendizaje, el cual consistió en estabilizar la latencia de cada ensayo. De esta forma, cada rata fue entrenada por 35 ensayos, los cuales fueron distribuidos en seis sesiones durante tres días. La primera sesión incluyó cinco ensayos, mientras que las otras cinco incluyeron seis ensayos cada una. La gráfica mostrada en la Figura V.19(a) presenta la latencia promedio de cada ensayo considerando el desempeño de los tres animales, mientras que la gráfica de la Figura V.19(b) muestra la latencia promedio de cada ensayo por sesión. En ambas gráficas es posible apreciar la estabilización de la latencia en los ensayos comprendidos dentro de la última sesión de entrenamiento.

La latencia promedio alcanzada al terminar el entrenamiento fue de 12 segundos, considerando que la rata que mejor aprendió la tarea logró llegar a la meta en sólo dos segundos. Durante las primeras cuatro sesiones, las ratas siguieron rutas indirectas de la localidad TD al bebedero, esto es, rutas que incluyen ciclos. A partir de la quinta sesión, algunos ensayos implicaron trayectorias directas aunque no necesariamente óptimas. En particular, la rata que mejor aprendió la tarea siguió un camino directo y óptimo en tres ensayos de la quinta sesión y en todos los ensayos de la sexta, mientras que los otros animales aún fallaron en dos y tres ensayos de la última sesión.

En síntesis, la latencia promedio de cada ensayo de entrenamiento fue de 122 segundos, por lo que la duración total promedio de la fase fue de una hora con 12 minutos por animal.

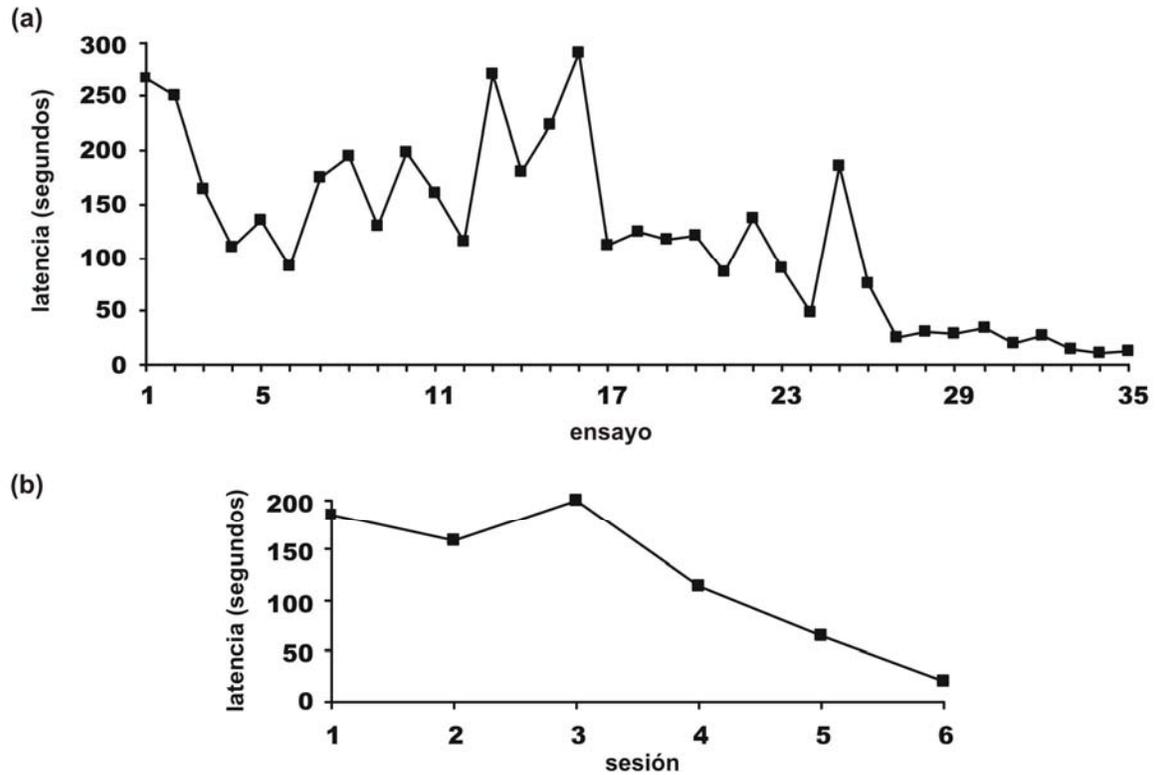


Figura V.19. (a) Latencia promedio de cada ensayo considerando el desempeño de tres ratas durante el entrenamiento en la tarea espacial desarrollada en el laberinto cíclico. (b) Latencia promedio de cada ensayo de entrenamiento por sesión, donde la sesión 1 incluye 5 ensayos, y las demás sesiones, 6. En (a) y en (b) es factible apreciar la estabilización de la latencia al finalizar el entrenamiento.

V.4.1.2. Fase de Prueba

Un día después de haber concluido el entrenamiento, las ratas fueron probadas durante 16 ensayos siendo liberadas desde cuatro localidades distintas del laberinto cíclico (D1, D2, D3, D4) en cuatro orientaciones diferentes (norte, sur, este, oeste). La Figura V.20 indica tales localidades de partida.

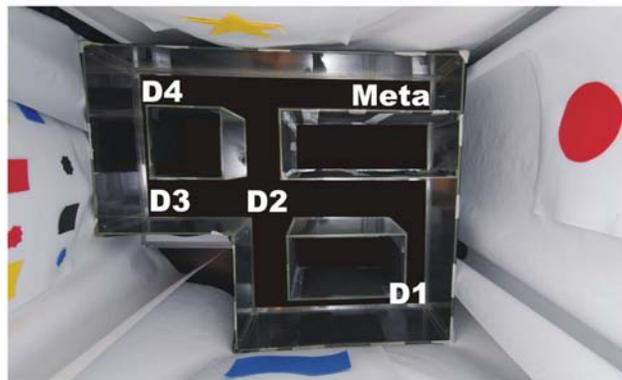


Figura V.20. Localidades de liberación de las ratas en los ensayos de prueba dentro del laberinto cíclico: D1, D2, D3 y D4. En cada localidad, las ratas son liberadas en cuatro direcciones: norte (90°), sur (270°), este (0°) y oeste (180°).

Como en cualquier ensayo de entrenamiento, en cada prueba la rata es extraída de su caja de acrílico y cubierta con una toalla; después es colocada en la localidad y orientación que correspondan, y el ensayo consiste en explorar libremente el laberinto hasta alcanzar la localidad meta y consumir el agua dulce. Al terminar la prueba, la rata es extraída del laberinto, cubierta nuevamente con una toalla y depositada en su caja.

Durante los ensayos de prueba se registraron tanto las latencias como las trayectorias que siguieron los animales para encontrar el depósito de agua. En particular, la rata que mejor aprendió la tarea reportó los resultados que presenta la Figura V.21 para el caso de la localidad D1, la Figura V.22 para el caso de la localidad D2, la Figura V.23 considerando la localidad D3, y la Figura V.24 para el caso de D4.

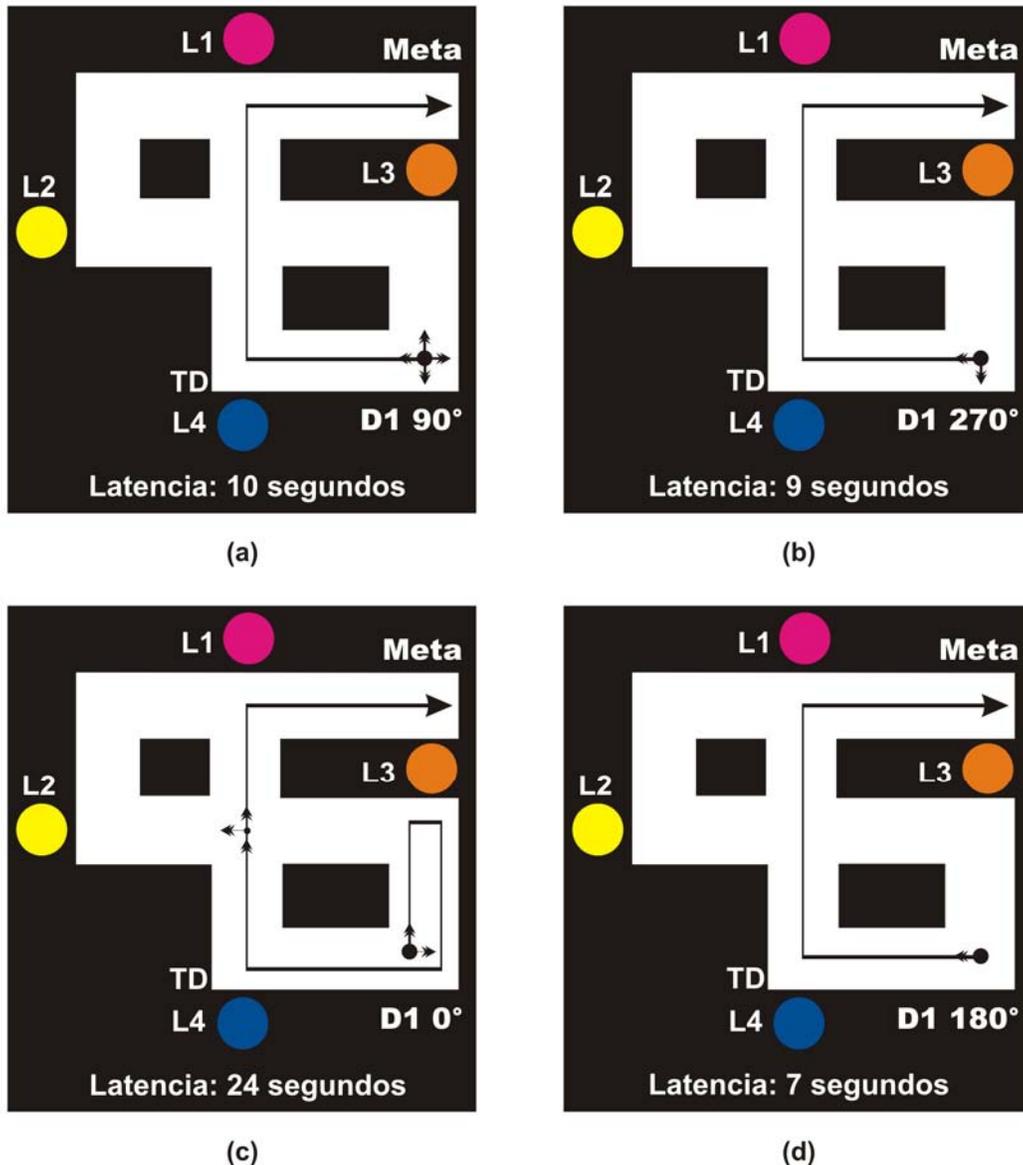


Figura V.21. La ruta seguida por la rata para encontrar el depósito de agua partiendo de la localidad D1 en las siguientes direcciones: (a) norte (90°), (b) sur (270°), (c) este (0°), y (d) oeste (180°). Las pequeñas flechas mostradas sobre las rutas indican giros de cabeza y/o de cuerpo ejecutados por el animal. Para cada ensayo de prueba se especifica su latencia en segundos.

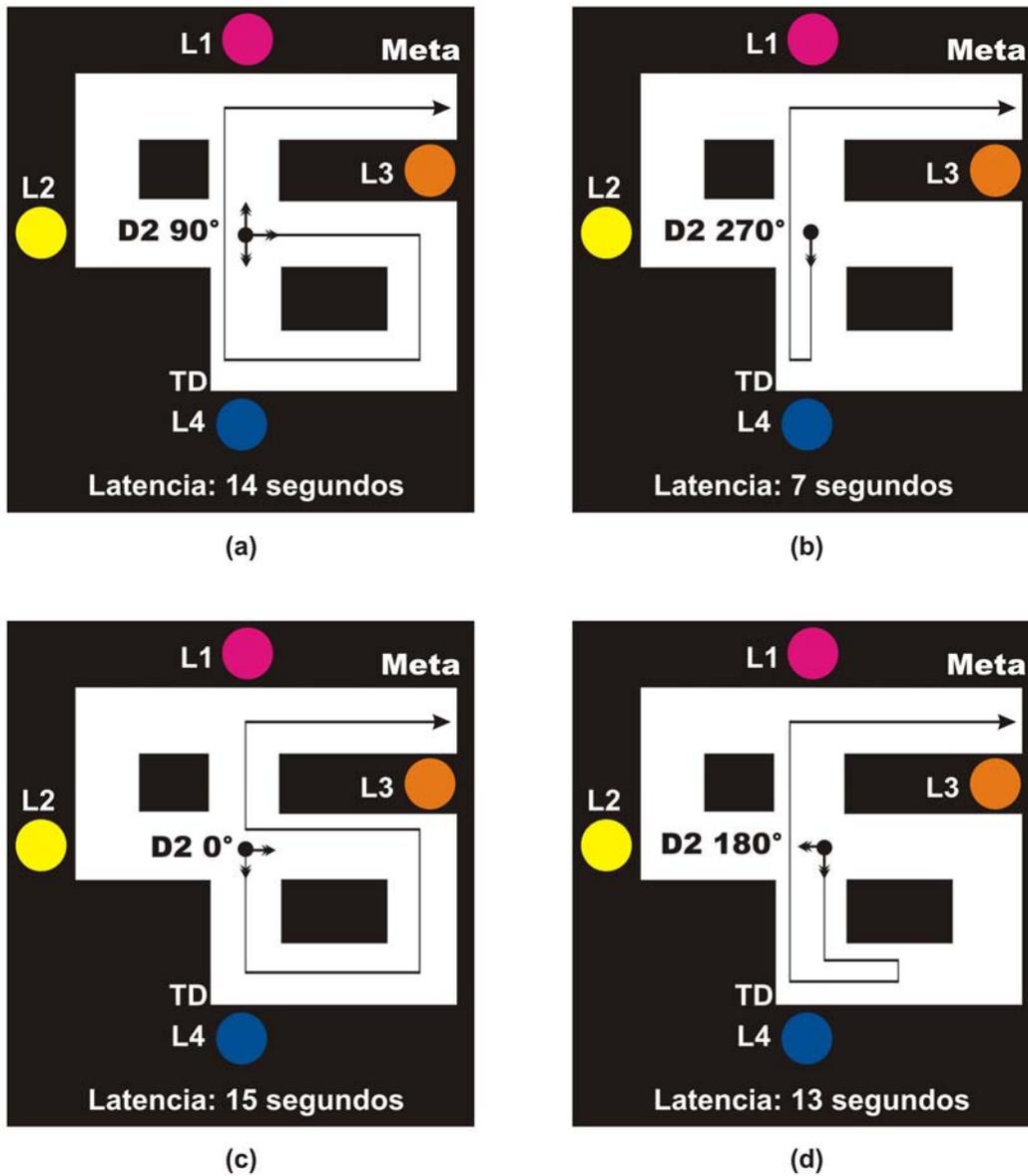


Figura V.22. La ruta seguida por la rata para encontrar el depósito de agua partiendo de la localidad D2 en las siguientes direcciones: (a) norte (90°), (b) sur (270°), (c) este (0°), y (d) oeste (180°). Las pequeñas flechas mostradas sobre las rutas indican giros de cabeza y/o de cuerpo ejecutados por el animal. Para cada ensayo de prueba se especifica su latencia en segundos.

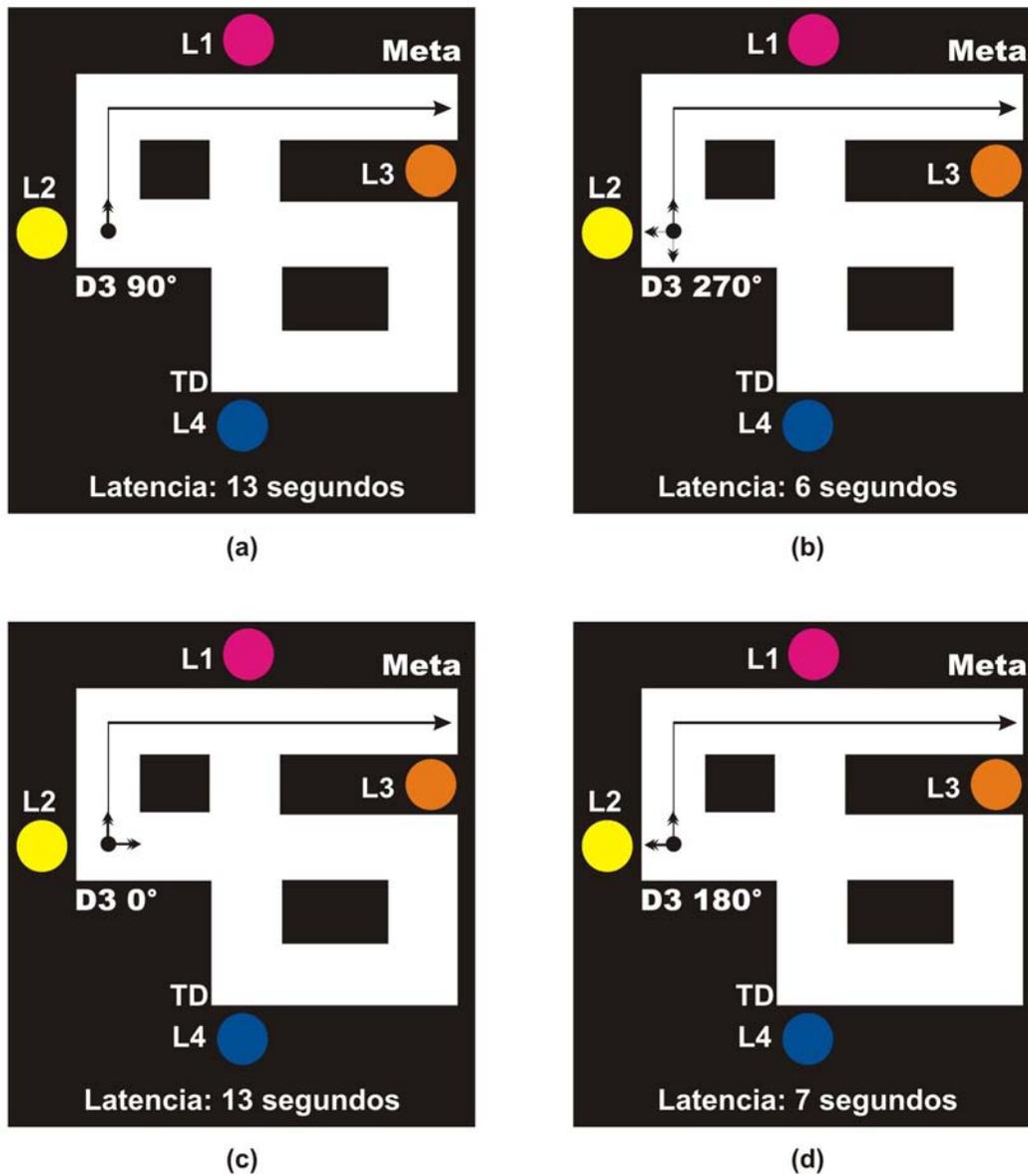


Figura V.23. La ruta seguida por la rata para encontrar el depósito de agua partiendo de la localidad D3 en las siguientes direcciones: (a) norte (90°), (b) sur (270°), (c) este (0°), y (d) oeste (180°). Las pequeñas flechas mostradas sobre las rutas indican giros de cabeza y/o de cuerpo ejecutados por el animal. Para cada ensayo de prueba se especifica su latencia en segundos.

gradiente de meta sugerido por Hull (1932), según el cual, la rata tiende a eliminar errores en las localidades de decisión del laberinto cercanas a la meta de manera más temprana que en aquellas encontradas al inicio del laberinto, considerando que el efecto del refuerzo es el máximo en la meta y que éste disminuye progresivamente conforme el animal se desplaza hacia atrás a través del laberinto.

V.4.2. El Laberinto Cíclico: Resultados Experimentales con Robots

Durante el experimento en el laberinto mostrado en la Figura V.18(a), el robot es colocado en la localidad identificada como TD y orientado a 90° (norte). Entonces comienza la fase de habituación, en la que no se ofrece recompensa alguna al robot mientras éste explora libremente el laberinto por un periodo de tiempo. Al concluir esta fase, se introduce una recompensa en la localidad HE que indica la Figura V.18(a), y el robot es entrenado desde el punto de partida TD para aprender la localidad correcta. Finalmente, el robot es probado para navegar hacia la meta desde diversas localidades y orientaciones. Los siguientes apartados describen las tres fases del experimento, y presentan los resultados derivados de cada una.

V.4.2.1. Fase de Habituación

El robot es liberado de la localidad TD indicada en la Figura V.18(a) estando orientado a 90° (norte). A partir de ahí, explora libremente el laberinto hasta alcanzar la localidad identificada como HE. Como se mencionó recientemente, no existe un reforzador en esta fase inicial del experimento, por lo que el robot no lleva a cabo un proceso de aprendizaje de localidades meta, y sus acciones están determinadas únicamente por curiosidad y aleatoriedad.

La duración de la exploración libre por parte del robot fue variable considerando tres ejecuciones del modelo. En promedio, el robot tardó 2 horas con 15 minutos en llegar a la localidad HE desde TD.

Como resultado de la fase de habituación, el robot construye el mapa cognitivo del laberinto conteniendo nodos para representar las localidades exploradas, y unidades Actor asociadas con las diversas direcciones en que éstas fueron exploradas. Así, el mapa generado por el robot en la mejor de las ejecuciones del modelo representa todas las localidades del laberinto en todas las direcciones posibles de exploración. Como se ilustra en el mapa de la Figura V.25, el robot utilizó 17 nodos para representar el ambiente en su totalidad y arcos bidireccionales para conectarlos.

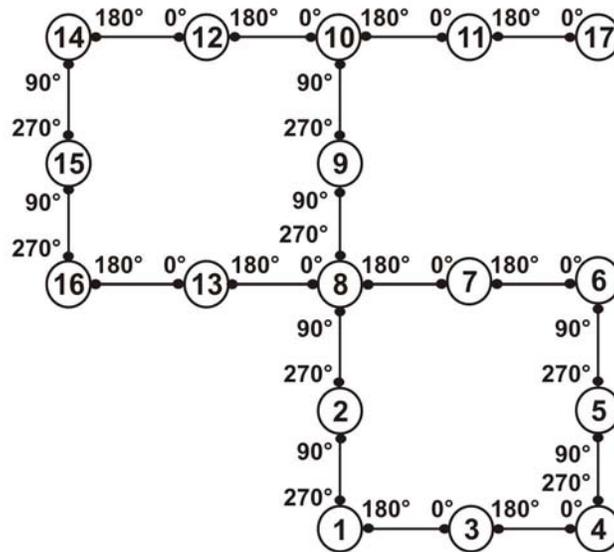


Figura V.25. El mapa cognitivo generado por el robot durante la fase de habituación de la tarea espacial en el laberinto cíclico. Los nodos se muestran numerados en orden de creación, y los arcos bidireccionales entre nodos están etiquetados con la dirección allocéntrica del robot cuando se desplazó de un lugar al siguiente.

V.4.2.2. Fase de Entrenamiento

Al concluir el proceso de habituación del robot, se introduce un reforzador en la localidad HE del laberinto (ver Figura V.18(a)). El robot es colocado en la localidad TD estando orientado a 90° (norte), y el entrenamiento inicia a partir del mapa cognitivo previamente construido.

Dado que no existen cambios en la configuración física del laberinto con respecto a la fase de habituación, la información visual que el robot percibe es la misma, así que los patrones de actividad registrados por las células de lugar del modelo son similares a los almacenados en el mapa cognitivo, lo que permite al robot reconocer las localidades que explora durante el entrenamiento.

En cada ensayo de esta fase, el robot explora el laberinto desde TD hasta encontrar la meta. Durante la exploración, marca los arcos que conectan los nodos explorados en la dirección correspondiente tantas veces como los visite. El modelo analiza el mapa al terminar el ensayo, y pueden ocurrir dos situaciones: (i) si el robot visitó un nodo más de una vez, la ruta de TD a la meta no es reforzada pues se trata de un camino indirecto que incluye ciclos; en cambio, (ii) si el robot visitó cada nodo sólo una vez, regresa al punto de partida siguiendo la misma ruta en dirección opuesta, y el proceso de refuerzo hacia atrás es aplicado a cada nodo del mapa que pertenezca a dicha trayectoria directa. Como ocurre normalmente, el refuerzo aplicado disminuye conforme aumenta la distancia entre el lugar representado por el nodo actual y la localidad meta. Concretamente, los nodos 14, 12, 10 y 11 en dirección 0° , y los nodos 8 y 9 en dirección 90° , obtienen un mayor refuerzo (ver Figura V.25).

Aún cuando durante la exploración de entrenamiento, la curiosidad dejó de ser uno de los factores que determinan las acciones del robot debido a que reconoce las localidades del laberinto, la aleatoriedad continúa presente provocando que el robot siga rutas indirectas o cíclicas hacia la meta, sobre todo en regiones alejadas de la localidad objetivo y mientras las expectativas de recompensa del robot no se incrementan lo suficiente para sobrepasar el ruido. Esto implicó ensayos de entrenamiento prolongados, así que con la finalidad de reducir la duración total de esta fase, decidimos ajustar los parámetros del modelo relacionados con el proceso de refuerzo hacia atrás de tal manera que un solo ensayo en que el robot siguiera una ruta directa de TD a la meta bastara para aprenderla.

En las tres ejecuciones del modelo se llevaron a cabo ocho ensayos de entrenamiento. Cada ensayo duró en promedio 9 minutos, y en el mejor de los casos, el robot logró explorar todas las posibles rutas directas de TD a la meta, las cuales son ilustradas por la Figura V.26.

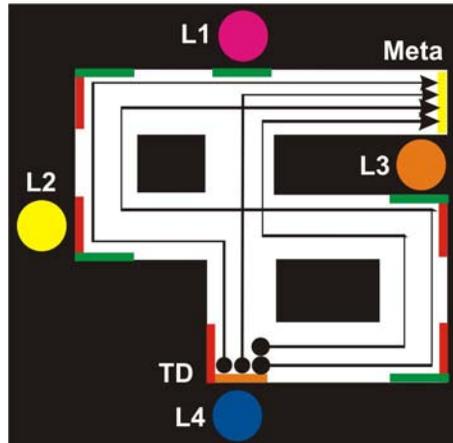


Figura V.26. Rutas directas exploradas por el robot de TD a la localidad meta durante la fase de entrenamiento del experimento en el laberinto cíclico.

V.4.2.3. Fase de Prueba

Al concluir el entrenamiento, ningún cambio físico fue inducido en el ambiente. Durante diversos ensayos de prueba, el robot fue colocado en distintas localidades de partida con diferentes orientaciones. La Figura V.27 ilustra las localidades de inicio utilizadas, D1, D2, D3 y D4, donde el robot parte en direcciones 90° (norte), 90° (norte), 0° (este) y 270° (sur), respectivamente.

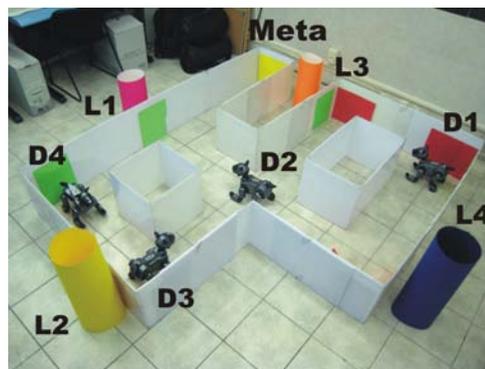


Figura V.27. Localidades de partida del robot en cuatro ensayos de prueba (D1, D2, D3 y D4), indicando su orientación en cada caso (90° , 90° , 0° y 270°).

Para cada posición de inicio, la Figura V.28 muestra la mejor ruta seguida por el robot para encontrar la meta, y el mapa del laberinto marcando los nodos que reconoce como pertenecientes a dicha ruta.

Gracias a la exploración exhaustiva efectuada durante la fase de habituación, el robot fue capaz de reconocer los lugares del laberinto previamente visitados en la misma dirección en todos los ensayos de prueba, como lo ilustra la Figura V.28, por lo que no requirió agregar nuevos nodos al mapa cognitivo. Este reconocimiento le permitió explotar la información almacenada en las unidades Actor de los nodos para alcanzar la localidad meta ejecutando las rotaciones asociadas con las máximas expectativas de recompensa.

Específicamente, en la prueba que inició en D3, por ejemplo (Figura V.28(c)), el robot reconoce la localidad representada por el nodo 16 y previamente explorada en dirección 0° , por lo que elige girar a la izquierda para orientarse a 90° , dirección asociada con la máxima expectativa de recompensa. Al llegar al fondo del pasillo, reconoce el nodo 14 en dirección 90° y decide rotar a la derecha motivado por la expectativa de recompensa que corresponde a la dirección 0° . Se desplaza en tal dirección, y reconoce los nodos 12 y 10. La expectativa de recompensa de la unidad Actor 0° del nodo 10 hace que el robot siga avanzando en vez de girar a la derecha, y alcance entonces la localidad meta.

La latencia de los ensayos de prueba ilustrados en la Figura V.28 para llegar a la meta desde cada localidad inicial fue de 150 segundos para el caso de D1, 80 segundos para D2, 130 segundos para D3, y 90 segundos para D4.

Finalmente, conviene aclarar que el hecho de que el robot haya logrado mantener un solo nodo para representar un mismo lugar desde diversas direcciones no implica que la respuesta de las células de lugar del modelo sea multidireccional, como sucede con la actividad de estas neuronas en el hipocampo de la rata. De hecho, cada nodo del mapa sigue estando asociado a un conjunto de unidades Actor que almacenan los patrones de actividad registrados por la colección de celdas de lugar en cada dirección de exploración de la localidad.

V.4.3. Análisis Conductual Comparativo

Esta sección presenta una comparación entre la conducta de la rata y la del robot durante la tarea espacial en el laberinto cíclico.

El experimento llevado a cabo con el robot implicó una fase inicial de habituación que le permitió mejorar notablemente su desempeño durante el entrenamiento y las pruebas al lograr reconocer las localidades del laberinto en múltiples direcciones. Aún cuando la tarea con ratas no incluyó dicha fase inicial, éstas se habituaron naturalmente al nuevo ambiente mientras fueron entrenadas. En efecto, durante los ensayos prolongados en las primeras sesiones de entrenamiento, los animales exploraron el laberinto en su totalidad avanzando, girando y regresando. Así lograron aprender la localidad meta en las últimas sesiones para poder desplazarse en forma rápida durante los ensayos de prueba partiendo de diversas localidades.

Aunque la cantidad de ensayos de entrenamiento del robot fue menor que la de las ratas (i.e., 8 ensayos del robot por 35 de cada rata), la duración total promedio del entrenamiento fue la misma (i.e., una hora con 12 minutos), pues un ensayo de la rata dura 2 minutos en promedio, mientras que en el caso del robot dura 9 minutos.

Como resultado del entrenamiento, la latencia de los ensayos se estabiliza en el caso de las ratas, y el robot logra aprender las diversas rutas directas que conducen a la meta desde la localidad fija designada.

Durante las pruebas, el robot fue capaz de reconocer los lugares del laberinto previamente visitados en la misma dirección, lo que le permitió explotar la información almacenada en las unidades Actor de los nodos del mapa cognitivo para alcanzar la localidad meta ejecutando las rotaciones asociadas con las máximas expectativas de recompensa. Las trayectorias seguidas por el robot desde las localidades D3 en dirección este (0°), y D4 en dirección sur (270°) coinciden con las ejecutadas por la rata que mejor aprendió la tarea. En las dos situaciones se trata de rutas directas y óptimas hacia la meta. Considerando el caso de la localidad de liberación D2 en dirección norte (90°), en cambio, la rata siguió un camino largo hacia el depósito de agua,

mientras que el robot reportó una trayectoria óptima a la meta. Finalmente, desde la localidad D1, la rata y el robot ejecutaron rutas distintas aunque directas y proporcionalmente de la misma longitud. Desde luego, la latencia de los ensayos de prueba del robot resultó mayor que la registrada en el caso de la rata, pues el robot emplea 10 segundos aproximadamente tan sólo en tomar las tres lecturas de las imágenes antes de ejecutar cada paso en su recorrido.

Los resultados derivados del experimento con ratas no revelan un desempeño eficiente en todas las situaciones probadas. De hecho, como se describió en la Sección V.4.1.2, la rata reportó casos de trayectorias indirectas que implicaron pasar dos veces por los mismos lugares aún después de haber aprendido la tarea.

Por otro lado, la conducta exitosa que manifestó el robot durante el experimento tampoco garantiza un desempeño eficiente en cualquier situación, ya que éste depende de un mapa cognitivo construido y mantenido de manera precisa durante las fases de habituación y entrenamiento de la tarea. En concreto, situaciones en las que el robot no explore el ambiente exhaustivamente durante la habituación, o no visite todas las posibles rutas directas del punto de partida fijo a la meta durante el entrenamiento, implicarán, de una parte, la adición de nuevos nodos al mapa afectando el reconocimiento de lugares durante las pruebas y la explotación de las máximas expectativas de recompensa que conducen a la meta de forma directa, y de otra, la navegación aleatoria del robot hasta reconocer un nodo del mapa que pertenezca a la ruta o rutas aprendidas. Además, si en los ensayos de prueba el robot parte de localidades muy alejadas de la meta, el valor de las expectativas de recompensa asociado a esos nodos del mapa puede ser inferior al nivel de ruido, dado el refuerzo decremental que el modelo aplica sobre las rutas que conducen al objetivo, por lo que el robot no necesariamente seguirá una ruta directa ni óptima. La Sección VI.I incluye una explicación más detallada al respecto.

VI. DISCUSIÓN Y CONCLUSIONES

La experimentación que se llevó a cabo con el robot en diversas tareas espaciales permitió demostrar que el modelo propuesto de cognición espacial y navegación en ratas le atribuye, en tiempo real, las siguientes capacidades:

- Generación de un mapa cognitivo topológico-métrico del ambiente durante la exploración del mismo.
- Aprendizaje y desaprendizaje de localidades meta a través de una arquitectura de refuerzo.
- Explotación del mapa cognitivo para encontrar la localidad objetivo desde cualquier punto de partida, ejecutando así navegación orientada a metas.
- Adaptación del mapa cognitivo para representar cambios físicos del ambiente inducidos de manera posterior a la exploración de entrenamiento.
- Reconocimiento de lugares en múltiples direcciones después de la exploración exhaustiva del ambiente.

Antes de concluir discutiendo los resultados derivados del proyecto de investigación, conviene destacar las diferencias entre la arquitectura de nuestro modelo y la del trabajo previo considerado como fuente de inspiración (Guazzelli et al., 1998), así como las mejoras propuestas al mismo.

La arquitectura de Guazzelli y colaboradores no se presenta organizada de manera modular, como en nuestro caso, sino que su funcionalidad se distribuye directamente entre las capas que la componen. La distinción primaria entre ambos modelos se relaciona con el proceso de aprendizaje que lleva a cabo el animal.

El modelo de Guazzelli y colaboradores implementa dos arquitecturas de aprendizaje del tipo Actor-Crítico: una para llevar a cabo la navegación directa hacia una meta perceptible (*taxon*), y la otra para posibilitar la navegación hacia una localidad meta aprendida (*locale*). La primera arquitectura de aprendizaje es referida como esquema motivacional egocéntrico, y la segunda como esquema motivacional allocéntrico.

El sistema de navegación *taxon* depende únicamente de información relacionada con affordances, así que aún cuando nuestro modelo incorpora la misma noción de affordances y coincide con el de los autores en la codificación de éstas dentro de esquemas perceptuales, su arquitectura incluye dos capas adicionales empleadas para producir estados de affordances que representan las distintas situaciones motrices experimentadas por el animal dentro de un ambiente y experimento dados. Dichos estados son mantenidos durante la ejecución completa del modelo, y el método de aprendizaje Actor-Crítico es utilizado para reforzar las affordances específicas dentro de los estados que al ser ejecutadas le permiten a la rata percibir la meta. Posteriormente, este refuerzo contribuye a la determinación de las decisiones del animal cuando experimenta situaciones motrices particulares. El gran problema involucrado en esta manera de implementar el método Actor-Crítico consiste en asumir que la rata experimentará únicamente situaciones motrices distintas mientras explora el ambiente, pues en caso de encontrar el mismo estado de affordances en dos contextos diferentes, fracasará al ejecutar la misma affordance reforzada en el contexto dentro del cual no se percibe la meta. A pesar de que el aprendizaje por refuerzo recae sobre los estados de affordances, en la arquitectura del modelo de los autores se observa el flujo de comunicación de la capa que produce los estados al esquema

motivacional egocéntrico, pero no al revés, así que no está clara la forma en que tal capa recibe la señal de refuerzo efectivo producida por el esquema egocéntrico.

Por otro lado, el sistema de navegación *locale* depende de información kinestésica y visual, y utiliza la gráfica del mundo que los autores construyen en otra de las capas del modelo. La arquitectura de aprendizaje Actor-Crítico implementada en el esquema motivacional alocéntrico refuerza direcciones específicas en los nodos de la gráfica que representan localidades desde las cuales se percibe la meta. El grave problema en este caso radica en asumir que el ambiente explorado por el animal incluye únicamente una localidad de decisión y que la meta es perceptible desde ese punto. De ocurrir esta situación, el nodo que representa la localidad de decisión será reforzado adecuadamente en la dirección de la meta; pero cuando se trate de una meta imperceptible a distancia, como típicamente sucede en los experimentos reales, la rata sólo podrá aprender a llegar a la meta desde el nodo anterior al que representa dicha localidad objetivo. A pesar de que el aprendizaje por refuerzo recae sobre los nodos de la gráfica del mundo, en la arquitectura del modelo de los autores se observa el flujo de comunicación de la gráfica al esquema motivacional alocéntrico, pero no al revés. En contraste, nuestra arquitectura define claramente el flujo de comunicación del módulo de representación de lugares al de aprendizaje para efectuar la predicción de recompensas, y del módulo de aprendizaje al de representación de lugares para actualizar el mapa cognitivo en consecuencia.

El proceso de aprendizaje propuesto en nuestro modelo complementa el uso de la arquitectura Actor-Crítico, que permite aprender y desaprender la localidad meta en su vecindad dentro del ambiente, con el método de propagar el refuerzo hacia atrás a través de los nodos de la trayectoria seguida por el animal durante cada ensayo del experimento. En este método, la adaptación de las expectativas de recompensa asociadas con las direcciones apropiadas en las diversas localidades de la trayectoria, hace posible el aprendizaje y desaprendizaje de rutas en ambientes que incluyen varios puntos de decisión y donde la meta no es perceptible a distancia. Una vez explorado y aprendido el ambiente por parte de la rata, la explotación de las máximas expectativas de recompensa almacenadas en los nodos del mapa cognitivo le permiten alcanzar la meta siguiendo una ruta directa desde cualquier localidad inicial.

En la arquitectura de Guazzelli y colaboradores, los dos esquemas motivacionales participan en la selección de la próxima acción a ejecutar por el animal, la cual también es determinada por las affordances actuales, por la curiosidad de la rata, y por una señal referida como estímulo relevante que no es más que la magnitud de la rotación que conduce a la meta visible. En nuestro modelo, el módulo de selección de acción recibe las máximas expectativas de recompensa producidas por el módulo de representación de lugares, el esquema perceptual de affordances generado por el módulo de procesamiento de affordances, y considera también la curiosidad del animal por las rotaciones aún no ejecutadas desde la localidad actual. Sin embargo, cuando las expectativas de recompensa no son significativas y la curiosidad se ha extinguido, las acciones de la rata son determinadas al azar a través de un esquema perceptual que representa una rotación aleatoria entre las affordances actuales.

En el modelo de los autores existe una entidad que lleva a cabo la funcionalidad de nuestro módulo de motivación, el cual se encarga de calcular el valor del hambre y de la recompensa inmediata que proporciona al animal la presencia de comida. No obstante, dicha entidad se comunica directamente con la capa que construye la gráfica del mundo, a diferencia de lo que ocurre en nuestra arquitectura, donde la recompensa inmediata es enviada al módulo de aprendizaje para predecir la recompensa futura.

Como extensión adicional al modelo propuesto por Guazzelli y colaboradores, el módulo de representación de lugares dentro de nuestra arquitectura implementa un proceso de adaptación en línea del mapa cognitivo ante situaciones en las que el ambiente cambia su configuración física durante la exploración por parte del animal. Al respecto, el modelo distingue entre dos situaciones de adaptación, la integración y la creación / eliminación de nodos del mapa, que pueden implicar la reasignación a nuevos nodos de la información relativa a las expectativas del robot por encontrar recompensas en localidades específicas.

Por lo que se refiere a los módulos de integración de trayectoria y de procesamiento de landmarks incluidos en nuestro modelo, la estructura interna de ambos coincide con la propuesta por Guazzelli y

colaboradores. Nuestra implementación particular se relaciona con la forma de representar el ambiente para actualizar la posición del punto de partida del animal cada vez que éste se mueve, y con la técnica para construir los esquemas perceptuales de landmarks a partir de la información de distancia y orientación hacia ellos.

En síntesis, es posible afirmar de acuerdo con el análisis efectuado del modelo propuesto por Guazzelli y colaboradores, que éste provee una implementación efectiva del sistema de navegación *taxon*, pero no del sistema *locale*, ya que aún cuando construye un mapa cognitivo del ambiente, no lo explota adecuadamente ni lo adapta para alcanzar la localidad meta desde cualquier punto de partida dentro de un laberinto simple, o incluso desde un mismo punto de partida dentro de laberintos complejos que incluyan más de una localidad de decisión.

Por último, en cuanto al contexto biológico que subyace al modelo de los autores, la funcionalidad del esquema motivacional aloentríco está relacionada con el estriado ventral, el cual influye directamente sobre el esquema de selección de acción. En nuestro modelo, el estriado no se comunica con el módulo de selección de acción, sino que es la corteza prelímbica quien le envía las máximas expectativas de recompensa que determinan la próxima acción de la rata. En ambos modelos, el hipocampo envía al estriado información acerca de la localidad actualmente explorada por el animal. No obstante, nuestro modelo sugiere que el estriado debiera influir al hipocampo a través del giro dentado enviando las expectativas de futura recompensa correspondientes a las acciones de la rata para posibilitar su desempeño apropiado durante la ejecución de tareas espaciales reforzadas.

El resto de este capítulo de la disertación presenta la discusión final concerniente a los resultados experimentales obtenidos del proyecto de investigación, a su contribución en los campos de la robótica y de las neurociencias, y a las líneas de investigación que serán abordadas como metas próximas.

VI.1. SOBRE LOS RESULTADOS EXPERIMENTALES

La implementación en el robot de la tarea espacial de discriminación inversa de O'Keefe en el laberinto tipo T y en el radial de 8 brazos reveló la similitud conductual entre el robot y las ratas. En promedio, las ratas sanas y las ratas con hipocampo lesionado tardan 20 ensayos de prueba en desaprender la opción previamente correcta y aprender la nueva. En el laberinto tipo T, los dos grupos de animales manifiestan un cambio abrupto en su elección del brazo incorrecto al brazo correcto, pero en el laberinto radial, las acciones de las ratas lesionadas, a diferencia de las ratas sanas, muestran un cambio sistemático y gradual del giro originalmente recompensado al nuevo giro que ofrece la recompensa. En ambos casos, O'Keefe argumentó que el cambio abrupto de girar a la izquierda a girar a la derecha en el laberinto T es un reflejo del momento en que la orientación del animal cruza la línea media en el laberinto radial, pasando del cuadrante de rotaciones negativas al cuadrante de rotaciones positivas.

El comportamiento recién descrito para el caso de las ratas es similar al obtenido con robots "sanos" y robots "lesionados". La diferencia ocurre en el criterio de desaprendizaje-aprendizaje, pues ambos "tipos de robots" tardan en promedio 12 ensayos en vez de 20. Sin embargo, eliminar esta diferencia es factible aplicando el ajuste apropiado sobre los parámetros de aprendizaje del modelo propuesto.

En la tarea de discriminación inversa, el criterio depende de la persistencia que implica la hipótesis empleada por el animal durante el aprendizaje. Las ratas sanas utilizan una hipótesis de lugares, esto es, consideran la información de las localidades del laberinto para soportar sus decisiones en los puntos de elección. Las ratas con lesiones en el hipocampo, en cambio, sólo pueden fundamentar sus acciones en una hipótesis de orientación que involucra el seguimiento de rutas o procedimientos, como concatenación de orientaciones a la izquierda o a la derecha, por ejemplo. Aún cuando la hipótesis de lugares es menos

persistente que la de orientación, dado que la ausencia inesperada de recompensa motiva la exploración del ambiente y la adaptación del mapa cognitivo, el criterio presentado por los dos grupos de ratas fue el mismo (20 ensayos). O'Keefe explicó este hecho en términos del sobreentrenamiento de los animales sanos en la tarea, el cual provoca que dejen de considerar la hipótesis de lugares y empiecen a utilizar la hipótesis de orientación como lo hacen las ratas lesionadas. Durante el entrenamiento, los robots “sanos” también utilizan ambas hipótesis. La hipótesis de lugares les permite construir el mapa cognitivo, definir una expectativa de recompensa en la localidad más cercana a la meta, e incrementar la expectativa de recompensa por acciones específicas a ejecutar en las localidades que conforman la ruta a seguir desde el punto de partida hacia la meta. Adicionalmente, la hipótesis de orientación es utilizada por los robots al ejecutar las rotaciones con el máximo valor de expectativa de recompensa en las localidades pertenecientes a la ruta que conduce a la meta cuando tales valores superan la curiosidad y el ruido.

De acuerdo con O'Keefe, la estrategia de aprendizaje de localidades empleada por las ratas sanas en el laberinto tipo T y en el radial de 8 brazos consideró información derivada de retroalimentación kinestésica y de la forma del laberinto en ausencia de señales visuales externas. En el caso del robot, únicamente la información kinestésica determinó la actividad conjunta de las células de lugar, la cual fue almacenada en el mapa cognitivo para representar las diversas localidades del laberinto.

Por otro lado, en la tarea espacial efectuada con el robot en el laberinto de múltiples Ts rodeado por landmarks, la actividad conjunta de las células de lugar almacenada en el mapa cognitivo quedó determinada tanto por retroalimentación kinestésica como por la información visual obtenida a partir de la percepción de los landmarks externos. Aún cuando el protocolo de esta tarea fue inspirado en el laberinto de agua de Morris, no es factible realizar una comparación apropiada entre los resultados obtenidos con robots y los reportados por Morris al experimentar con ratas sanas, dadas las variaciones importantes implementadas al experimento original. No obstante, fue posible confirmar en el caso de los robots “sanos”, como lo hizo Morris con las ratas sanas, que para encontrar una meta “escondida” en el laberinto independientemente de la localidad de partida, los robots requieren explotar su mapa cognitivo reconociendo y utilizando las expectativas de recompensa que les hacen navegar hacia la localidad objetivo exitosamente.

La experimentación que se llevó a cabo en el mismo laberinto de múltiples Ts rodeado por landmarks introduciendo diversos cambios físicos después de la fase de entrenamiento, permitió verificar que el robot, en tiempo real, es capaz de adaptar el mapa cognitivo agregando y/o eliminando nodos para reflejar los cambios percibidos. No obstante, en cuanto a la navegación hacia la meta, el tipo de cambio inducido en la configuración de landmarks puede implicar tres escenarios:

- (i) que el robot no reconozca las localidades representadas en el mapa cognitivo actual y por ende no explore la información relacionada con las expectativas de recompensa correspondientes, que reexplora el ambiente mientras construye un nuevo mapa soportando sus decisiones únicamente sobre la base de la curiosidad y el ruido, y que no logre encontrar la localidad meta;
- (ii) el mismo escenario (i), pero encontrando finalmente la localidad meta; o
- (iii) que el robot reconozca algunas de las localidades representadas en el mapa cognitivo actual, y utilice las expectativas de recompensa asociadas a ellas para poder encontrar la meta, como ocurrió en los casos mostrados dentro de la Sección V.3.

Por último, la exploración exhaustiva efectuada durante la fase de habituación en el laberinto cíclico permitió al robot representar en el mapa cognitivo todas las localidades visitadas en todas las orientaciones posibles, agregar la información relacionada con expectativas de recompensa durante el entrenamiento, reconocer los lugares en todos los ensayos de prueba sin agregar nuevos nodos al mapa, y explotar la información almacenada en las unidades Actor de los nodos para alcanzar la localidad meta ejecutando las rotaciones asociadas con las máximas expectativas de recompensa. Sin embargo, la eliminación de la multiplicidad en la representación de localidades dentro del mapa no implica que el modelo propuesto provea un reconocimiento de lugares multidireccional, pues la respuesta de las células de lugar depende de la dirección actual del robot, así que cada nodo del mapa sigue estando asociado a un conjunto de unidades

Actor que almacenan los patrones de actividad registrados por la colección de celdas de lugar en cada dirección de exploración de la localidad.

Tanto en la fase de habituación del experimento en el laberinto cíclico, como en la de entrenamiento y en la de prueba, la Sección V.4 de la disertación reportó casos exitosos, en donde el robot logra explorar exhaustivamente el ambiente en un tiempo razonablemente corto, aprender todas las posibles rutas directas del punto de partida fijo a la meta, mostrar una mejoría absoluta en el proceso de reconocimiento de lugares en múltiples direcciones, y encontrar la localidad objetivo desde diversas posiciones de inicio en poco tiempo siguiendo rutas directas e incluso óptimas. No obstante, situaciones inconvenientes también pueden llegar a ocurrir.

Concretamente, si durante la habituación el robot no explora el ambiente en su totalidad, en fases posteriores agregará nuevos nodos al mapa para representar localidades desconocidas o perspectivas desconocidas de una localidad previamente explorada en una dirección particular, lo cual le impedirá reconocer los lugares visitados en los ensayos de prueba, y por tanto, explotar las expectativas de recompensa almacenadas en el mapa para llegar a la meta de forma directa y en poco tiempo. Por otro lado, si durante el entrenamiento el robot no visita todas las posibles rutas directas del punto de partida fijo a la meta, éstas no serán reforzadas por el modelo, así que durante las pruebas desde diversas localidades de inicio, el robot navegará al azar hasta reconocer un nodo del mapa que pertenezca a la ruta o rutas aprendidas y a partir de ahí explotar las máximas expectativas de recompensa que permiten alcanzar la localidad meta. En tal caso, no es probable que las trayectorias del robot en el encuentro de la meta sean óptimas. Finalmente, si en los ensayos de prueba el robot parte de localidades muy alejadas de la meta, el valor de las expectativas de recompensa asociado a esos nodos del mapa lejanos al nodo meta puede ser inferior al nivel de ruido, dado el refuerzo decremental que el modelo aplica sobre las rutas que conducen al objetivo. Ante esta situación, el robot no necesariamente seguirá una ruta directa y mucho menos óptima, sino que navegará al azar hasta encontrar un nodo con un buen nivel de expectativa de recompensa que lo haga llegar a la meta a partir de ahí.

En cuanto a los resultados derivados del experimento con ratas en el laberinto cíclico, éstos no revelaron un desempeño eficiente en todas las situaciones probadas. De hecho, la rata reportó casos de trayectorias indirectas que implicaron pasar dos veces por los mismos lugares aún después de haber aprendido la tarea.

Además de los escenarios recién descritos, como se discute más adelante en la Sección VI.3, prevalecen las dificultades que involucran los laberintos más grandes y complejos que el pequeño ambiente cíclico utilizado durante la experimentación con el robot: el manejo de colores para resolver las opciones de desplazamiento en cada instante de tiempo (affordances), la dependencia del buen cálculo de affordances durante todo el experimento para crear y mantener un mapa cognitivo preciso y confiable, el tiempo que implica habituar al robot antes de comenzar la tarea y entrenarlo de manera que visite todas las posibles rutas directas al objetivo.

VI.2. SOBRE LA CONTRIBUCIÓN DE LA INVESTIGACIÓN

La contribución de la investigación debe ser discutida desde dos perspectivas: la robótica y la neurocientífica.

La revisión literaria sobre SLAM sintetizada en la Sección II.2 de la disertación, permitió identificar que la mayor parte de la investigación actual versa sobre la propuesta de algoritmos de mapeo y localización computacionalmente más eficientes, de soluciones a los problemas de asociación de datos y ambigüedad perceptual para el reconocimiento preciso de lugares, y de representaciones ambientales de distintos tipos. Sin embargo, pocos esfuerzos han sido documentados en relación al uso de tales representaciones

espaciales para navegar directamente hacia localidades meta designadas. Aún más, no hemos encontrado reportes de algoritmos propuestos para manejar el proceso de desaprendizaje de localidades meta previamente aprendidas ante cambios en el ambiente.

Desde la perspectiva robótica, entonces, la contribución del presente trabajo consiste en abordar la inequidad entre mapeo y explotación del mapa que existe en la literatura sobre SLAM, así como la falta de investigación sobre desaprendizaje, tomando inspiración de tales habilidades en las ratas para proponer un modelo de cognición espacial y navegación que habilita a un robot, en tiempo real, para explorar un ambiente controlado, generar un mapa topológico-métrico del mismo, ejecutar navegación orientada a metas, y adaptar el mapa para representar cambios físicos del ambiente. En efecto, la predicción y explotación de expectativas de recompensa asociadas a localidades incluidas en la representación espacial, permiten al robot tanto aprender como desaprender localidades meta y encontrar rutas directas para alcanzarlas desde cualquier punto de partida en el ambiente.

El interés particular del trabajo radica en entender los mecanismos que subyacen al proceso de cognición espacial en las ratas, con el fin de incorporar datos fisiológicos relativos a dicho proceso en una arquitectura robótica que sea evaluada desde una perspectiva conductual a través de la comparación con resultados biológicos. En particular, como lo describe el Capítulo V de la disertación, dicha evaluación comparativa se efectuó para el caso de la clásica tarea de discriminación inversa implementada por O'Keefe (1983), y para la tarea espacial en el laberinto cíclico diseñada por nosotros. De esta forma, se considera que la investigación contribuye a la neurociencia experimental ofreciendo el sistema propuesto como una herramienta para probar con robots nuevas hipótesis que pudieran extender el conocimiento actual sobre aprendizaje y memoria en ratas.

Concretamente, la contribución se relaciona con la actividad de los neurocientíficos durante su labor experimental. En general, para especificar el experimento a realizar, el neurocientífico integra el conocimiento anatómico y fisiológico que se refiera al fenómeno que estudia. Después, efectúa el experimento por algún tiempo, obtiene resultados y plantea hipótesis. Las hipótesis lo conducen a prever nuevos resultados, así que el próximo experimento que realice involucrará tales hipótesis con el afán de ratificar o rectificar sus predicciones. Dado que el modelo propuesto captura algunas propiedades de las estructuras del cerebro de la rata involucradas en aprendizaje y memoria, puede ser empleado como una plataforma de experimentación robótica que permita al neurocientífico probar sus hipótesis mediante la simulación de nuevos experimentos. En caso de obtener resultados interesantes de la experimentación con el robot, el siguiente experimento que efectúe con animales habrá sido sugerido por la herramienta, y los resultados reales, difieran o no de los artificiales, servirán al neurocientífico para plantear nuevas hipótesis y/o para ajustar las hipótesis incrustadas en el modelo computacional. El modelo podrá incluir las innovaciones o adaptaciones, y el ciclo reiniciará para sugerir el próximo experimento a realizar.

Una contribución adicional del modelo propuesto, en su estado actual, consiste en sugerir cómo el núcleo estriado, involucrado en el sistema de recompensas del cerebro de la rata, debiera influir al hipocampo, asociado con el sistema de representación espacial, a través del giro dentado, considerando descubrimientos recientes que indican que las ratas con lesiones en el giro dentado están severamente impedidas en su habilidad de actualizar información espacial esencial para guiar comportamientos orientados a metas. Específicamente, como las unidades Actor incluidas en el modelo corresponden a las posibles orientaciones que la rata pueda adoptar estando en cualquier lugar dado, y las expectativas de recompensa asociadas a tales orientaciones por el estriado son actualizadas durante el entrenamiento en una tarea espacial reforzada para permitir la posterior navegación directa hacia la meta, se sugiere que las unidades Actor pueden localizarse en el giro dentado. Así, el estriado debiera enviar al hipocampo las expectativas de futura recompensa correspondientes a las acciones del animal para ser aprovechadas por el giro dentado en el desempeño apropiado del animal durante la ejecución de tareas espaciales reforzadas.

Finalmente, dentro del campo de la robótica, muchos aspectos deben ser abordados antes de situar el sistema propuesto dentro de una escala de desempeño o compararlo con los resultados obtenidos por enfoques particulares de SLAM; y desde la perspectiva biológica, el modelo básico de cognición espacial y

navegación en ratas debe ser sofisticado incorporando conocimiento adicional sobre aprendizaje y memoria, con la finalidad de habilitarlo para proveer retroalimentación valiosa a la neurociencia experimental. Así, los aspectos a tratar como trabajo futuro tanto en el ámbito de la robótica como de la biología, quedan descritos en el próximo apartado.

VI.3. SOBRE LAS LÍNEAS FUTURAS

Si se considera como meta futura del trabajo abordar problemas robóticos artificiales mediante la explotación de modelos biológicos, el sistema propuesto habrá de ser mejorado en los siguientes aspectos:

- La determinación de affordances u opciones de desplazamiento del robot en cada instante de tiempo. Actualmente, la interfaz entre el modelo y el ambiente depende del reconocimiento de colores para determinar las posibles affordances. En los diversos laberintos probados, papeles de colores han sido colocados sobre la pared en el fondo de los pasillos, con el propósito de definir si el robot puede seguir avanzado y/o puede rotar hacia alguna dirección. Antes de comenzar un experimento en un laberinto particular, la cámara local del robot es empleada para tomar lectura de la cantidad de píxeles de colores percibidos por el robot desde distintos lugares y orientaciones. De esa forma, se definen rangos de píxeles para cada posible rotación del robot estando orientado hacia cada dirección. Aunque este proceso implica bastante tiempo, ha resultado manejable para el caso de laberintos simples y controlados, pero dejará de serlo cuando se intente experimentar con ambientes más grandes que incluyan más puntos de decisión. De hecho, aún en los laberintos simples, cuando ocurren accidentes relacionados con la mecánica del robot como un resbalón, su orientación se vuelve imprecisa y por ende la lectura de las cantidades de píxeles de colores también, provocando entonces la determinación incorrecta de las affordances. Si esto sucede, el experimento debe ser interrumpido y reiniciado, ya que el no hacerlo ocasiona que el mapa cognitivo se distorsione dada su dependencia en los cambios de affordances para crear y mantener los nodos. En síntesis, las opciones de desplazamiento del robot deben ser definidas por un mecanismo alternativo escalable a ambientes más complejos, el cual seguramente implicará el uso de sensores de proximidad.
- El sistema perceptual. La estimación de las distancias y orientaciones relativas de los landmarks perceptibles por el robot en cada instante de tiempo, depende de un procesamiento visual similar al descrito en el punto anterior, en donde cualquier interferencia o ruido en la lectura de los píxeles de colores provoca el cálculo erróneo de la información visual de entrada a las células de lugar del modelo, afectando, en consecuencia, el reconocimiento de lugares previamente explorados. Se requiere diseñar, por tanto, un sistema perceptual alternativo capaz de manejar retos reales, de tal manera que permita al modelo garantizar la robustez en el reconocimiento de lugares por parte del robot.
- La eficiencia computacional del módulo de procesamiento de landmarks. En su estado actual, el modelo agrega una capa de neuronas para generar y detectar patrones de información visual derivados de cada landmark disponible dentro del ambiente. Posteriormente, otra capa produce y detecta patrones de información visual derivados de la combinación de todos los landmarks visibles en cada instante de tiempo. El mayor número de landmarks probado hasta ahora ha sido cuatro, y el tiempo de procesamiento no se ha convertido en un problema. Sin embargo, ambientes más grandes requerirán más landmarks para que el robot perciba alguna(s) desde la mayoría de los lugares, y no dependa únicamente de información kinestésica para representarlos y reconocerlos. Más landmarks incrementarán seguramente el tiempo de procesamiento, así que el modelo habrá de ser ajustado para incluir una sola capa de neuronas que produzca patrones de información visual derivados directamente de la integración de todos los landmarks visibles, sin necesidad de capas intermedias.

- La inicialización y retroalimentación kinestésica del integrador de trayectoria. En la implementación actual del modelo, la técnica de remapeo dinámico para integrar la trayectoria del robot implica el uso de una matriz con tamaño proporcional al ambiente, en donde las coordenadas del punto de partida están predeterminadas y son actualizadas en cada instante de tiempo considerando las rotaciones y traslaciones del robot. Esta información kinestésica no proviene de sensores de movimiento del robot, sino que el modelo lleva un registro actualizado de la dirección del robot a partir de las salidas motrices del módulo de selección de acción producidas en cada momento. De esta forma, si el robot está orientado a 90°, por ejemplo, y se le indica girar 45° a la izquierda, pero en realidad gira un poco más o menos, el modelo considera que efectivamente rotó 45° y que su nueva dirección alocéntrica es 135°. Puede ocurrir, entonces, que los errores se acumulen, y en consecuencia, que las affordances se calculen de forma incorrecta y que el mapa se distorsione. Introducir al sistema propuesto el uso de sensores de movimiento que puedan informar acerca de la magnitud real de la rotación y desplazamiento del robot, permitirá implementar correcciones en la orientación del mismo que prevengan errores posteriores en su desempeño.
- La optimización de rutas para alcanzar el objetivo. La arquitectura de aprendizaje implementada en el modelo en combinación con la estrategia del refuerzo hacia atrás, permiten al robot tanto aprender la localidad meta como las rutas que conducen a ella desde un punto de partida fijo. Cuando el robot parte de una localidad distinta, navegará al azar hasta reconocer un lugar perteneciente a alguna de las rutas aprendidas. Esa posible navegación aleatoria inicial puede ocasionar que la ruta seguida por el robot para alcanzar el objetivo no sea óptima, aún cuando disponga de un mapa completo del ambiente construido en la fase de habituación y mantenido durante el entrenamiento. La estrategia actual para optimizar las trayectorias del robot está relacionada con el protocolo experimental de entrenamiento, en el que se persigue que el robot explore todas las posibles rutas directas desde el punto de partida fijo a la meta. Sin embargo, no es factible que esto ocurra siempre, así que la mejora de este aspecto bien puede implicar el uso de alguno de los algoritmos computacionales clásicos que permiten determinar la ruta más corta entre dos puntos de un grafo, considerando el costo que corresponda en tiempo de procesamiento.
- El tiempo de exploración durante la habituación del robot. Durante la fase de habituación, el robot explora el ambiente de forma exhaustiva motivado por curiosidad y azar en un tiempo variable, pudiendo visitar las localidades varias veces en la misma o en diferentes direcciones. Desde luego, la duración de la fase se extiende conforme la complejidad del laberinto aumenta. Actualmente, el nivel de curiosidad que maneja el modelo es bajo, así que será conveniente aumentarlo para dar preferencia a los caminos aún no visitados por el robot, reduciendo entonces el tiempo de exploración.
- La escalabilidad a ambientes de mayor extensión y complejidad. Los ambientes en que el modelo ha sido demostrado con éxito son laberintos artificiales pequeños con claves visuales bien definidas. Los aspectos discutidos en los puntos anteriores evidencian las dificultades involucradas en el manejo de ambientes más grandes y menos controlados, donde los pasillos no estén necesariamente orientados en alguna de las ocho direcciones que maneja el modelo, los landmarks no sean artificialmente dispuestos por el experimentalista sino identificados naturalmente por el robot, o incluso donde no existan pasillos sino un campo delimitado pero abierto en el que la noción de affordances pierda sentido.

Por otra parte, dado el objetivo de contribuir a la neurociencia experimental aportando retroalimentación valiosa a través del modelo propuesto, éste habrá de ser enriquecido analizando e implementando los siguientes aspectos:

- La respuesta multidireccional de las células de lugar. En la versión actual del modelo, cuando el robot está situado en una localidad particular, la actividad conjunta de las células de lugar es direccional; es decir, resulta distinta cada vez que el robot modifica su orientación, ya que la entrada visual puede cambiar. De esta forma, considerando que el robot es capaz de orientarse en ocho direcciones, una localidad del ambiente es representada en el mapa a través de ocho diferentes patrones de actividad de las células de lugar, y cada vez que el robot recibe la misma entrada visual estando orientado hacia una de las ocho

direcciones, reconoce el patrón de actividad de las celdas de lugar asociado con tal dirección. En el caso de las ratas, por el contrario, el mismo grupo de celdas de lugar responde cuando el animal se orienta hacia diversas direcciones estando en una misma localidad física. Este hecho será estudiado para determinar la estrategia de modelado que permita adaptar la respuesta de las células de lugar de unidireccional a multidireccional.

- La distinción entre las células de lugar encontradas en las áreas CA3 y CA1 del hipocampo. El módulo actual de representación de lugares no distingue entre las células de ambas regiones, aún cuando se sabe que la dinámica de grupo de CA3 y CA1 presenta diferencias. Concretamente, estudios recientes han examinado las respuestas de las células de lugar encontradas en ambas áreas ante cambios en el contexto ambiental de las ratas. Los resultados evidencian que CA3 es capaz de ejecutar transformaciones no lineales de los patrones de entrada sensorial, mientras que CA1 representa los cambios en la entrada de manera lineal (Guzowski et al., 2004). Adicionalmente, se ha analizado la actividad grupal de las regiones al observar cómo las células generan representaciones de cuartos con elementos espaciales comunes. En CA3, subconjuntos distintos de células se activaron en cada cuarto, mientras que en CA1, las poblaciones activadas se traslaparon. Ante la exposición posterior a un nuevo cuarto, la actividad conjunta se desarrolló más despacio en CA3 que en CA1, sugiriendo que las representaciones emergen de forma independiente (Leutgeb et al., 2004).
- La asimilación de la información kinestésica y la posterior integración de la información visual en la respuesta de las células de lugar. En el estado actual del modelo, las celdas de lugar reciben e integran las entradas kinestésica y visual al mismo tiempo. Sólo en caso de ausencia de landmarks externos en el ambiente, la actividad de las células queda determinada únicamente por la información procedente del integrador de trayectoria. No obstante, se sabe que las propiedades de disparo de las células de lugar son determinadas de manera inicial por información kinestésica independientemente de claves externas, cuando no existen landmarks familiares, o en la oscuridad sin alguna referencia espacial previa (McNaughton et al., 1996). De esta forma, las ratas suelen depender más en los mecanismos de integración de trayectoria que en landmarks hasta que aprenden que los landmarks constituyen referencias direccionales estables (Knierim et al., 1996).
- El modelado de células de dirección de cabeza en el integrador de trayectoria. Se ha sugerido que un componente esencial del sistema integrador de trayectoria es una población de neuronas que registra la orientación de la cabeza del animal, principalmente con base en señales de movimiento angular. Tales células de dirección de cabeza han sido encontradas en un conjunto de estructuras cerebrales dispersas pero interconectadas, que incluyen la corteza parietal posterior, la corteza retrosplenial, el presubiculum dorsal y el tálamo anterior (McNaughton et al., 1996), y disparan siempre en la misma dirección angular relativa sin importar el ambiente o la localidad en que el animal se encuentre. Este sistema de dirección de cabeza constituye una especie de brújula interna para la rata, y es la fuente de información direccional para las células de lugar del hipocampo (Etienne y Jeffery, 2004). El trabajo futuro de modelar las células de dirección de cabeza como parte de la arquitectura propuesta, resulta fundamental para que el robot sea capaz de determinar su orientación actual, y puede estar relacionado con la meta de adaptar la respuesta de las células de lugar de unidireccional a multidireccional.
- El ajuste de los niveles de proyección entre las capas neuronales del modelo. En esta versión básica del modelo, cada neurona en una capa está conectada aleatoriamente al 50% de las neuronas en la siguiente capa. Específicamente, esta situación ocurre en los módulos de integración de trayectoria, de procesamiento de landmarks, y de representación de lugares. Como extensión futura, este alto nivel de proyección será ajustado al nivel conocido o supuesto por los neurocientíficos, concretamente, entre la corteza parietal posterior y la corteza retrosplenial, entre la corteza parietal posterior y la corteza entorhinal, entre la corteza retrosplenial y el hipocampo, y entre la corteza entorhinal y el hipocampo.

VII. APÉNDICE A: VALOR DE PARÁMETROS

Este apéndice incluye dos tablas que presentan los valores de los principales parámetros utilizados en las ecuaciones del modelo de cognición espacial y navegación en ratas sanas (Tabla VII.A.I), y en las ecuaciones de la versión adaptada del modelo para el caso de ratas con hipocampo lesionado (Tabla VII.A.II).

Tabla VII.A.I. Valor de los principales parámetros utilizados en la implementación del modelo de cognición espacial y navegación en ratas sanas.

PARÁMETRO	DESCRIPCIÓN	VALOR
nPS	Cantidad de neuronas en cualquier esquema perceptual lineal.	80
d	Varianza de cualquier distribución Gaussiana.	3
N	Cantidad de neuronas en cualquier capa detectora de patrones (i.e., PIFDL, LFDL, LL, PCL).	400
nN	Cantidad de vecindarios dentro de una capa detectora de patrones (i.e., PIFDL, LFDL, LL, PCL).	5
m	Cantidad de neuronas dentro de cada vecindario.	80
h	Altura de cualquier distribución de Gauss en el esquema perceptual de affordances (APS), en cualquier esquema perceptual de landmark (LPS), y en el esquema perceptual de remapeo dinámico (DR).	1
hR	Altura de la distribución Gaussiana centrada en una posición aleatoria dentro del esquema perceptual aleatorio (RPS).	0.04
hC	Altura de cualquier distribución de Gauss en el esquema perceptual de curiosidad (CPS).	0.05
hE	Altura de la distribución de Gauss en el esquema perceptual de expectativa global de máxima recompensa (EMR).	≥ 1
d_{max}	Máximo valor del instinto de hambre empleado por el módulo de motivación.	20
α_d	Factor intrínseco de hambre utilizado por el módulo de motivación.	0.003
a	Reducción del instinto de hambre después de ingerir comida.	0.2
b	Incremento del instinto de hambre después de percibir un incentivo.	0.15
n	Dimensión de la máscara M empleada por DRL para codificar la dirección aloéctrica opuesta de la cabeza de la rata.	3
α	Tasa de aprendizaje utilizada en cualquier aplicación de la regla de Hebb.	0.001
γ	Factor de descuento usado por el módulo de aprendizaje para calcular el refuerzo efectivo.	0.85
β	Tasa de aprendizaje empleada por el módulo de aprendizaje para actualizar los pesos de conexión entre PCL y AC, y entre las unidades Actor y los nodos del mapa.	0.041
χ	Factor de actualización aplicado por el módulo de aprendizaje a los rastros de elegibilidad de las conexiones entre PCL y AC.	0.3
τ	Factor de actualización aplicado por el módulo de aprendizaje al rastro de elegibilidad de la conexión entre el nodo activo en el mapa y la unidad Actor asociada con la dirección de la rata.	0.1
λ	Factor de decrecimiento aplicado por el módulo de aprendizaje a todos los rastros de elegibilidad.	0.8

Tabla VII.A.II. Valor de los principales parámetros utilizados en la implementación del modelo de navegación para ratas con hipocampo lesionado.

PARÁMETRO	DESCRIPCIÓN	VALOR
nPS	Cantidad de neuronas en cualquier esquema perceptual lineal.	80
d	Varianza de cualquier distribución Gaussiana.	3
N	Cantidad de neuronas en la capa detectora de patrones de affordances AFDL.	400
nN	Cantidad de vecindarios dentro de la capa detectora de patrones de affordances AFDL.	5
m	Cantidad de neuronas dentro de cada vecindario.	80
h	Altura de cualquier distribución de Gauss en el esquema perceptual de affordances (APS).	1
$hR1$	Altura de la distribución Gaussiana centrada en una posición aleatoria dentro del esquema perceptual que representa la señal de ruido durante el entrenamiento (RPS).	0.03
$hR2$	Altura de la distribución Gaussiana centrada en una posición aleatoria dentro del esquema perceptual que representa la señal de ruido durante la tarea inversa (RPS).	0.001
hE	Altura de la distribución de Gauss en el esquema perceptual de expectativa global de máxima recompensa (EMR).	≥ 1
d_{max}	Máximo valor del instinto de hambre empleado por el módulo de motivación.	20
α_d	Factor intrínseco de hambre utilizado por el módulo de motivación.	0.003
a	Reducción del instinto de hambre después de ingerir comida.	0.2
b	Incremento del instinto de hambre después de percibir un incentivo.	0.15
α	Tasa de aprendizaje utilizada en la aplicación de la regla de Hebb.	0.001
γ	Factor de descuento usado por el módulo de aprendizaje para calcular el refuerzo efectivo.	1.10
β	Tasa de aprendizaje empleada por el módulo de aprendizaje para actualizar los pesos globales asociados a los estados de affordances y los pesos correspondientes a cada una de las rotaciones comprendidas dentro de los estados de affordances.	0.10
χ	Factor de actualización aplicado por el módulo de aprendizaje a los rastros de elegibilidad del estado de affordances activo y del estado de affordances que corresponde a la localidad en que el animal tomó su penúltima decisión.	0.3
τ	Factor de actualización aplicado por el módulo de aprendizaje a los rastros de elegibilidad que corresponden: (i) a la rotación ejecutada por la rata dentro del estado de affordances activo, (ii) a la rotación ejecutada dentro del estado de affordances anterior, y (iii) a la rotación adyacente a la rotación ejecutada dentro del estado de affordances anterior.	0.2
λ	Factor de decrecimiento aplicado por el módulo de aprendizaje a todos los rastros de elegibilidad.	0.9

VIII. GLOSARIO

Affordances de movimiento. Posibles rotaciones que la rata puede ejecutar de -180° a $+180^\circ$ en intervalos de 45° a partir de su localización y dirección de cabeza actuales.

Aprendizaje de Hebb. Cuando una capa de neuronas se proyecta sobre otra, esta regla de aprendizaje permite asegurar que cada vez que un patrón de actividad similar se presente en la capa fuente, el mismo conjunto de unidades se active en la capa destino, con el fin de posibilitar la detección de patrones de información.

Campo de lugar. Área física, continua y compacta donde la rata se localiza provocando la respuesta a una tasa elevada de una célula de lugar del hipocampo.

Células de lugar. Células piramidales encontradas en las áreas CA3 y CA1 del hipocampo que muestran una alta correlación entre su tasa de disparo y la localización del animal en el ambiente.

Cognición espacial. Proceso de codificación y almacenamiento de información espacial que permite que la orientación hacia los objetos y el movimiento a través del espacio sean exitosos.

Discriminación. Habilidad para responder de manera diferencial ante objetos o lugares en el ambiente. La tarea de discriminación comúnmente utilizada consiste en la presentación simultánea de dos opciones, recompensando al animal al responder a una de ellas pero no a la otra.

Discriminación inversa. Problema que consiste en desaprender una hipótesis previamente correcta y aprender una nueva. Durante una tarea de discriminación inversa, la opción previamente positiva se convierte en negativa y viceversa después de que el animal aprende a discriminar entre dos estímulos o lugares eligiendo consistentemente el positivo y evitando el negativo.

Esquema perceptual. Arreglo lineal que almacena una o más distribuciones de Gauss cuyas posiciones medias dentro del arreglo corresponden a rotaciones relativas específicas entre -180° y $+180^\circ$ en intervalos de 45° .

Expectativas de recompensa. Valor numérico que representa la expectativa de la rata de obtener una recompensa al orientarse hacia una dirección particular estando en la localidad actual.

Hipótesis de lugares. Estrategia del animal para resolver un problema de discriminación mediante el aprendizaje de las localidades que ofrecen recompensa.

Hipótesis de orientación. Estrategia del animal para resolver un problema de discriminación mediante el aprendizaje del procedimiento a seguir para obtener la recompensa.

Información kinestésica. Señales de retroalimentación interna procedentes de los músculos de un individuo durante su locomoción en el espacio. En el modelo propuesto, la información kinestésica incluye las magnitudes de la rotación y de la traslación ejecutadas por el animal en cualquier instante de tiempo.

Integración de trayectoria. Proceso a través del cual un sujeto en movimiento es capaz de mantener un registro de su posición en relación al lugar de partida. El término refleja la adición sucesiva de pequeños incrementos de movimiento dentro de una representación continuamente actualizada de la dirección y la distancia recorrida desde el punto de partida.

Landmarks. Señales o claves visuales que controlan la navegación de la rata cuando ésta las percibe como estables dentro del ambiente.

Latencia. Duración en segundos de cualquier ensayo ejecutado por el animal durante el desarrollo de un experimento dado.

Locale. Estrategia de aprendizaje basada en una hipótesis de lugares.

Mapa cognitivo. Representación espacial interna que almacena los patrones de actividad global de la población de células de lugar del hipocampo correspondientes a las diversas ubicaciones experimentadas por el animal dentro de un ambiente familiar.

Taxon. Estrategia de aprendizaje dirigida por la percepción de la recompensa o basada en una hipótesis de orientación.

IX. REFERENCIAS

- Arbib, M. A. 1987. *Brains, Machines and Mathematics*. Springer-Verlag.
- Arbib, M. A. 1989. *The Metaphorical Brain 2: Neural Networks and Beyond*. Wiley.
- Arbib, M. A. and Lieblich, I. "Motivational learning of spatial behavior." In Metzler, J. (Ed.), *Systems Neuroscience*, Academic Press, New York, pp. 221-239., 1977.
- Arleo, A., Smeraldi, F., Gerstner, W. 2004. Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. *IEEE Transactions on Neural Networks* 15 (3): 639-652.
- Barrera, A., Weitzenfeld, A. 2006. 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)*, Pisa, Italy.
- Barrera, A., Weitzenfeld, A. 2007. Rat-inspired Model of Robot Target Learning and Place Recognition, In *Proceedings of the 15th Mediterranean Conference on Control and Automation (MED)*, Athens, Greece.
- Bar-Shalom, Y., Fortmann, T. E. 1988. *Tracking and Data Association*. Academic Press.
- Barto, A. G. "Adaptive critics and the basal ganglia." In Houk, C., Davis, J. L., Beiser, D. (Eds.), *Models of information processing in the basal ganglia*, MIT Press, Cambridge, MA, pp. 215-232, 1995.
- Bosse, M., Newman, P., Leonard, J., Teller, S. 2004. SLAM in large-scale cyclic environments using the Atlas Framework. *International Journal on Robotics Research* 23 (12):1113-1139.
- Burgess, N., O'Keefe, J. "Spatial Models of the Hippocampus." In Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks*, 2nd Edition, MIT press, Cambridge, MA, 2002.
- Burgess, N., Recce, M., O'Keefe, J. 1994. A model of hippocampal function. *Neural Networks* 7 (6-7): 1065-1081.
- Cheatwood, J. L., Reep, R. L., Corwin, J. V. 2003. The associative striatum: cortical and thalamic projections to the dorsocentral striatum in rats. *Brain Research* 968 (1):1 - 14.
- Cho, J., Sharp, P. 2001. Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behavioral Neuroscience* 115 (1): 3-25.
- Cooper, B., Mizumori, S. 1999. Retrosplenial cortex inactivation selectively impairs navigation in darkness. *Neuroreport* 10 (3): 625-630.
- Corwin, J. V., Reep, R. L. 1998. Rodent posterior parietal cortex as a component of a cortical network mediating directed spatial attention. *Psychobiology* 26 (2): 87 - 102.
- Dempster, A. P., Laird, A. N., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39 (1): 1-38.
- Dominey, P. F., M. A. Arbib. 1992. A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex* 2: 153-175.
- Durrant-Whyte, H., Rye D., Nebot, E. "Localization of Autonomous Guided Vehicles." In Hirzinger, G., Giraldo, G. (Eds.), in *Proceedings of the 8th International Symposium on Robotics Research*, SpringerVerlag, New York, pp. 613 - 625, 1995.

- Etienne A. S., Boulens V., Maurer R., Rowe T., Siegrist C. 2000. A brief view of known landmarks reorientates path integration in hamsters. *Naturwissenschaften* 87 (11): 494–498.
- Etienne, A. S. 1998. Mammalian Navigation, Neural Models and Biorobotics. *Connection Science* 10 (3-4): 271–289.
- Etienne, A. S., Jeffery, K. J. 2004. Path integration in mammals. *Hippocampus* 14 (2): 180–192.
- Folkesson, J., Christensen, H. 2004. Graphical SLAM - A self-correcting map. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA.
- Franz, M. O., Schölkopf, B., Mallot, H. A., Bühlhoff, H. 1998. Learning view graphs for robot navigation. *Autonomous Robots* 5: 111-125.
- Frese, U. 2006. A discussion of Simultaneous Localization and Mapping. *Autonomous Robots* 20: 25-42.
- Gaussier, P., Revel, A., Banquet, J. P., Babeau, V. 2002. From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biological Cybernetics* 86: 15-28.
- Gibson, J. J. 1966. *The senses considered as perceptual systems*. Houghton Mifflin, Boston, MA.
- Grace, A., Floresco, S., Goto, Y., Lodge, D. 2007. Regulation of firing dopaminergic neurons and control of goal-directed behaviors. *Trends in Neurosciences* 30 (5): 220-227.
- Granon, S., Poucet, B. 2000. Involvement of the rat prefrontal cortex in cognitive functions: A central role for the prelimbic area. *Psychobiology* 28 (2): 229-237.
- Guazzelli, A., Corbacho, F. J., Bota, M. and Arbib, M. A. 1998. Affordances, motivation, and the world graph theory. *Adaptive Behavior* 6 (3-4): 435-471.
- Guivant, J., Nebot, E., Nieto, J., Masson, F. 2004. Navigation and mapping in large unstructured environments. *The International Journal of Robotics Research* 23 (4): 449-472.
- Guzowski, J., Knierim, J. Moser, E. 2004. Ensemble Dynamics of Hippocampal Regions CA3 and CA1. *Neuron* 44: 581-584.
- Hähnel, D., Burgard, W., Wegbreit, B. and Thrun, S. 2003. Towards lazy data association in SLAM. In *Proceedings of the 11th International Symposium of Robotics Research (ISRR)*, Sienna, Italy.
- Hebb, D. O. 1949. *The organization of behavior: a neuropsychological theory*. Wiley-Interscience, New York.
- Hernández-Rabaza, V., Barcia, J., Llorens-Martín, M., Trejo, J., Canales, J. 2007. Spared place and object-place learning but limited spatial working memory capacity in rats with selective lesions of the dentate gyrus. *Brain Research Bulletin* 72 (4-6): 315-323.
- Hollerman, J., Tremblay, L., Schultz, W. 2000. Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Progress in Brain Research* 126: 193–215.
- Hollup, S. A., Kjelstrup, K. G, Hoff, J., Moser, M. and Moser, E. I. 2001 (1). Impaired Recognition of the Goal Location during Spatial Navigation in Rats with Hippocampal Lesions. *The Journal of Neuroscience* 21 (12):4505–4513.
- Hollup, S. A., Molden, S., Donnett, J. G., Moser, M. and Moser, E. I. 2001 (2). Place fields of rat hippocampal pyramidal cells and spatial learning in the watermaze. *European Journal of Neuroscience* – 13: 1197-1208.
- Houk, J. C., Adams, J. L., and Barto, A. G. “A model of how the basal ganglia generate and use neural signals that predict reinforcement.” In Houk, J. C., Davis, J. L., Beiser, D. G. (Eds.), *Models of information processing in the basal ganglia*, MIT Press, Cambridge, MA, pp. 249-270, 1995.
- Hull, C. L. 1932. The goal gradient hypothesis and maze learning. *Psychological Review* 39: 25-43.
- Jeffery K. J. 1998. Learning of landmark stability and instability by hippocampal place cells. *Neuropharmacology* 37: 677–687.
- Jeffery, K. J., O'Keefe, J. M. 1999. Learned interaction of visual and idiothetic cues in the control of place field orientation. *Experimental Brain Research* – 127: 151-161.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, Journal of Basic Engineering* 82: 35-45.

- Kelley, A. 2004. Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neuroscience and Biobehavioral Reviews* 27 (8): 765-776.
- Kesner, R., Novak, J. 1982. Serial position curve in rats: role of the hippocampus. *Science* 218: 173–175.
- Knierim J. J., Kudrimoti H. S., McNaughton B. L. 1996. Neuronal mechanisms underlying the interaction between visual landmarks and path integration in the rat. *International Journal of Neural Systems* 7 (2): 213–218.
- Kuipers, B., Byun, Y. 1991. A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations. *Journal of Robotics and Autonomous Systems* 8: 47-63.
- Kuipers, B., Modayil, J., Beeson, P., MacMahon, M., Savelli, F. 2004. Local metrical and global topological maps in the Hybrid Spatial Semantic Hierarchy. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA.
- Lee, I., Kesner, R. 2003. Time-Dependent Relationship between the Dorsal Hippocampus and the Prefrontal Cortex in Spatial Memory. *The Journal of Neuroscience* 23 (4):1517–1523.
- Leutgeb S, Leutgeb JK, Treves A, Moser MB, Moser EI. 2004. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* 305(5688):1295 – 1298.
- Maaswinkel H, Whishaw I. Q. 1999. Homing with locale, taxon, and dead reckoning strategies by foraging rats: sensory hierarchy in spatial navigation. *Behavioural Brain Research* 99 (2): 143–152.
- McNaughton B. L., Barnes C. A., Gerrard J. L., Gothard K., Jung M. W., Knierim J. J., Kudrimoti H., Qin Y., Skaggs W. E., Suster M., Weaver K. L. 1996. Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *Journal of Experimental Biology* 199: 173–185.
- McNaughton B. L., Knierim J. J., Wilson, M. A. “Vector encoding and the vestibular foundations of spatial cognition.” In Gazzaniga, M. (Ed.), *The cognitive neurosciences*, MIT Press, Boston, pp. 585-595, 1994 (1).
- McNaughton, B., Mizumori, S., Barnes, C., Leonard, B., Marquis, M., Green, E. 1994 (2). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cerebral Cortex* 4: 27-39.
- Milford, M., Wyeth, G., Prasser, D. 2006. RatSLAM on the Edge: Revealing a Coherent Representation from an Overloaded Rat Brain. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, pp. 4060 – 4065.
- Mittelstaedt, M., Mittelstaedt, H. “Homing by path integration in a mammal.” In *Avian Navigation*, Papi, F., Wallraff, H. G., Springer Verlag, Berlin, pp. 290-297, 1982.
- Montemerlo, M., Thrun, S. 2003. Simultaneous localization and mapping with unknown data association using FastSLAM. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1985-1991.
- Morris, R. G. M. “An Attempt to Dissociate Spatial-Mapping and Working-Memory Theories of Hippocampal Function.” In W. Seifert (Ed.), *Neurobiology of the Hippocampus*. New York, Academic Press, pp. 405 – 432, 1983.
- Morris, R. G. M. 1981. Spatial localization does not require the presence of local cues. *Learning and Motivation* 12: 239 – 260.
- Movarec, H. P., Elfes, A. 1985. High resolution maps from wide angle sonar. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 116-121.
- Newman, P., Cole, D., Ho, K. 2006. Outdoor SLAM using visual appearance and laser ranging. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, Florida.
- O’Keefe, J. “Spatial memory within and without the hippocampal system.” In *Neurobiology of the Hippocampus*. W. Seifert (Ed.), Academic Press, New York, pp. 375 – 403, 1983.
- O’Keefe, J., Conway, D. H. 1978. Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research* 31: 573–590.
- O’Keefe, J., Dostrovsky, J. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Research*, 34(1): 171 – 175.
- O’Keefe, J., Nadel, L. 1978. *The hippocampus as a cognitive map*. Oxford University Press.

- Parron, C., Save, E. 2004. Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Experimental Brain Research* 159 (3): 349-359.
- Poucet, B. 1993. Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychological Review* 100 (2): 163-182.
- Quirk, G. J., Muller, R. U., Kubie, J. L. 1990. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience* 10 (6): 2008-2017.
- Redish, A. 1997. Beyond the cognitive map. Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Redish, A. and Touretzky, D. 1997. Cognitive maps beyond the hippocampus. *Hippocampus* 7 (1): 15-35.
- Risold, P., Thompson, R., Swanson, L. 1997. The structural organization of connections between hypothalamus and cerebral cortex. *Brain Research Reviews* 24 (2-3): 197-254.
- Roberts, W. A. 1998. Principles of animal cognition. McGraw Hill, USA, pp. 201 – 230.
- Savelli, F., Kuipers, B. 2004. Loop-closing and planarity in topological map-building. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1511-1517.
- Schultz, W., Tremblay, L., Hollerman, J. 1998. Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology* 37 (4-5): 421-429.
- Schultz, W., Tremblay, L., Hollerman, J. 2000. Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex* 10 (3): 272-283.
- Thrun, S., Bücken, A., Burgard, W., Fox, D., Fröhlingshaus, T., Hennig, D., Hofmann, T., Krell, M. and Schmidt, T. "Map Learning and High-Speed Navigation in RHINO." In Kortenkamp, D., Bonasso, R. P., and Murphy, R. (Eds.), *AI-based Mobile Robots: Case Studies of Successful Robot Systems*, MIT Press, 1998.
- Toates, F. 1986. Motivational systems. Cambridge University Press. New York.
- Tolman, E. 1948. Cognitive maps in rats and men. *Psychological Review* 55: 189-208.
- Ulrich, I., Nourbakhsh, I. 2000. Appearance-based place recognition for topological localization. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA, pp. 1023-1029.
- Weitzenfeld, A., Arbib, M., Alexander, A. 2002. *The Neural Simulation Language*. MIT Press.
- Zivkovic, Z., Bakker, B., Kröse, B. 2005. Hierarchical map building using visual landmarks and geometric constraints. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, Canada, pp. 7-12.