UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE ELECTRÓNICA

MECANISMO DE TOMA DE DECISIONES BIOINSPIRADO APLICADO COMO CONTROLADOR DE UN AGENTE AUTÓNOMO.

Tesis de Grado presentada por

Cristóbal Jesús Nettle Vacher.

como requisito parcial para optar al título de

Ingeniero Civil Electrónico

y al grado de

Magíster en Ciencias de la Ingeniería Electrónica

Profesor Guía
Dr. María José Escobar Silva

Valparaíso, 2016.

TÍTULO DE LA TESIS:

MECANISMO DE TOMA DE DECISIONES BIOINSPIRADO APLICADO COMO CONTROLADOR DE UN AGENTE AUTÓNOMO

AUTOR:

Cristóbal Jesús Nettle Vacher.

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el título de Ingeniero Civil Electrónico y el grado de Magíster en Ingeniería Electrónica de la Universidad Técnica Federico Santa María.

Dr. María José Escobar Silva _____

Dr. Milan S. Derpich Musa _____

Dr. Miguel Torres Torriti _____

Valparaíso, septiembre de 2016.

AGRADECIMIENTOS

A través de este espacio, presento mis agradecimientos a cada una de las personas que han tenido participación durante mi extenso periodo universitario. A lo largo del tiempo, a través de distintas etapas, he contado con un permanente apoyo.

En especial, agradezco la circunstancia de haber conocido a Josefa, dirigiendo nuestras vidas hacia haber formado nuestra familia, recibiendo el amor de Isabella. Agradezco el gran apoyo que me ha entregado Josefa, que me ha entregado mi familia, padres y hermanas, y la familia de Josefa.

Agradezco a María José pues leerá esto. Aparte, le agradezco por toda la paciencia que ha tenido conmigo, porque desde su personalidad amable, activa y preocupada me ha apoyado de forma constante, tanto en mi desarrollo profesional como personal.

Le agradezco a todas las personas con las que he compartido proyectos, ya que sin duda ese espacio constituye gran parte de mi aprendizaje: desde los más inútiles – probablemente sean los desarrollados con Antonio y León–, hasta los proyectos más complejos. Al respecto, le agradezco al Centro de Robótica la oportunidad de otorgar un espacio sobre el cual los estudiantes pueden materializar sus propios proyectos y aspiraciones.

A cada una de las personas que he nombrado espero haber podido transmitirle mis sinceros agradecimientos. Si es que no he tenido ese efecto, ojalá puedan leer esta sección de nuevo.

RESUMEN

El presente trabajo considera la extensión de un modelo de lazos cortico-ganglios basales (CBG), conjunto de estructuras corticales y subcorticales relacionadas con la toma de decisiones, a través de la integración de efectos asociados al nivel de dopamina tónica (tipo D1). La dopamina (DA), neurotransmisor asociado con procesos de aprendizaje y memoria, se ha relacionado con efectos en el comportamiento con respecto a la razón entre exploración y explotación. El modelo resultante considera características como la consideración de múltiples lazos paralelos – considerando decisiones en múltiples niveles –, reglas de plasticidad sináptica que describen un aprendizaje dopaminérgico basado en recompensas, y la modulación en los procesos de selección sobre la tendencia a la exploración de nuevas opciones, frente a la explotación de conocimiento previamente adquirido. Para evaluar el comportamiento del modelo con respecto a cambios en los niveles de DA, se simula la ejecución de una tarea de selección forzada de dos opciones, considerando aprendizaje entre selecciones. Los datos obtenidos durante los procesos de selección en la realización de esta tarea demuestran variaciones en el comportamiento, en términos de cuánto se promueve la exploración de nuevas opciones en contra de la explotación de la información aprendida, al modificar los niveles de DA tónica. A pesar de esta modificación sobre el comportamiento y el desempeño del modelo, las pruebas realizadas predicen que las señales internas de aprendizaje no se ven modificadas ante variaciones en los niveles de DA.

En conjunto, con el fín de evaluar la aplicabilidad del modelo propuesto como mecanismo de toma de decisiones, y en base a la importancia de la regulación entre exploración y explotación en una plataforma robótica, se describe la estructura de un controlador diseñado para enfrentar una tarea de supervivencia de dos recursos, aplicado sobre el robot MODI (MODular Intelligence). Durante la realización de esta tarea, el robot MODI debe aprender en tiempo real cuáles son las acciones que le permiten aumentar su esperanza de vida. Mediante simulaciones, se prueba que el modelo es utilizable como mecanismo de toma de decisiones, y que variaciones en los niveles de dopamina tónica modifican las habilidades de supervivencia del robot. Los datos obtenidos sugieren la existencia de un nivel de DA tónica constante tal que maximiza la esperanza de vida alcanzada por el robot.

Palabras Clave: Robótica bio-inspirada, Mecanismos de toma de decisiones, Ganglios basales, Lazos cortico-ganglios basales, Dopamina.

ABSTRACT

The present work extends a cortico-basal ganglia (CBG) loops model, a set of cortical and sub-cortical structures related to decision-making processes, through the incorporation of effects associated to type-D1 tonic dopamine levels inside the basal ganglia. Dopamine (DA), neurotransmitter associated to learning and memory related processes, has also been related to behabioral modulations of the trade-off between exploitation and exploration. The resulting model presents multiple parallel loops – considering multiple decision-making at different levels –, synaptic plasticity rules that describes reward-based dopaminergic learning, and the modulation of the selection processes to promote exploring new options against exploiting acquired knowledge. To test the behavioral changes on the proposed model in relation to the tonic DA level, a two-choice forced selection task is simulated, considering learning between every selection. Obtained data from the process of selections during the performance of this task effectively shows a modulation of the exploitation-exploration trade-off, just modifying the tonic DA level. Despite the modulation of the behavior (and, in consequence, to the performance), data predicts that changes in the tonic DA levels are transparent with respect to internal signals related to the learning processes. This means that the learning inside CBG loops is produced independently of variations of the tonic dopamine level.

Additionally, in order to test the feasibility of using the CBG loops as a decision-making mechanism, and considering that the exploitation-exploration trade-off is essential for a robotic platform, a robot controller is proposed. The controller is used to deal with a two-resources survival task, applied into a MODI (MODular Intelligence) robot. During the performance of this task, the MODI robot has to learn on-line which options are the ones that expands its expected lifetime. Performed simulations shows that the CBG loops model can be applied as a decision-making mechanism, while changes in the tonic DA level modulates the robots survival skills. Obtained data suggests that there is a constant tonic DA level such that the expected lifetime is the highest.

Keywords: Bio-inspired robotics, Decision-Making mechanisms, Basal ganglia, Cortico-basal ganglia loops, Dopamine.

ABREVIACIONES

Lista de siglas y abreviaciones

CBG	: cortico-ganglios basales.
DA	: dopamina.
Ctx	: corteza.
GPe	: globo pálido externo, perteneciente a los ganglios basales.
GPi	: globo pálido interno, perteneciente a los ganglios basales.
MSNs	: Neuronas Espinosas Medias.
SNc	: Sustancia negra pars compacta, perteneciente a los ganglios basales.
SNr	: Sustancia negra pars reticulata, perteneciente a los ganglios basales.
STN	: núcleo subtalámico, perteneciente a los ganglios basales.
STR	: estriado, perteneciente a los ganglios basales.
Th	: tálamo.
sp	: espigas (*spikes*).

CONTENIDO

Capítulo 1

INTRODUCCIÓN

1.1. Motivación y Definición del problema

Los objetivos que han llevado a un gran número de investigadores a lo largo del tiempo a establecer modelos implementables del cerebro, especialmente humano, son amplios y con un claro impacto. El establecimiento de estos modelos permite definir de forma concreta la naturaleza de desórdenes mentales, con miras a generar soluciones específicas sin ser basadas únicamente en resultados empíricos; permite realizar experimentación sin la necesidad de intervenir especímenes en el proceso; y genera avances en la comprensión del funcionamiento neuronal, esencial para establecer paradigmas adecuados a un nivel de sociedad [30]. Además, son modelos aplicables en el desarrollo de tecnologías bioinspiradas.

Este campo se ha abordado principalmente definiendo modelos para características específicas, que pueden ser integrados en modelos a gran escala. De esta forma, se han propuesto modelos que cada vez incluyen un mayor número de características concordantes con los procesos biológicos. Un ejemplo de esta evolución es la toma de decisiones: diversos autores sugieren relaciones entre los ganglios basales y la selección de acciones motoras (e.g. [54], [33]), posteriormente se definen las relaciones entre los núcleos que componen los ganglios basales (e.g. [35], [25]). Sobre esta base, en [29] se define un modelo capaz de realizar una selección entre dos opciones, que luego es extendido considerando de múltiples decisiones simultáneas y aprendizaje basado en recompensas [20]. Los avances al respecto han permitido definir de forma paulatina modelos a gran escala, por ejemplo [11], donde se utiliza un modelo de ganglios basales para la toma de decisiones definido en [52].

Paralelamente, se han desarrollado modelos sobre la influencia de estados emocionales en la toma de decisiones. Esta influencia se ha descrito como una ventaja evolutiva, generando un desempeño superior, por ejemplo, al evaluar el privilegiar exploración versus explotación [40]. La ansiedad, estado emocional, se ha relacionado tanto en humanos como en primates con niveles anormales de dopamina, neurotransmisor que actúa como señal de aprendizaje que afecta la plasticidad sináptica y la memoria [24]. Específicamente, se ha correlacionado niveles bajos de dopamina en estructuras pertenecientes a los ganglios basales en pacientes con trastornos de ansiedad [57], produciendo además cambios en la razón entre exploración y explotación [22]. Para un agente autónomo, la exploración es esencial para aprender y mejorar su comportamiento en relación a algún resultado esperado, sin embargo, expone al agente a una situación desconocida, potencialmente riesgosa, por lo que debe existir un

trade-off adecuado entre ésta y la explotación [4].

En base a la influencia positiva de las emociones con respecto al proceso de aprendizaje, interacción con el ambiente e interacción con humanos, es que teorías en el campo de la psicología indican que su integración puede ser un gran beneficio para un agente robótico autónomo [32]. En [1], se concluye que el uso de emociones puede ser aplicado para acelerar el aprendizaje y mejorar la toma de decisiones, abriendo la posibilidad de desarrollar máquinas más inteligentes, incluyendo toma de decisiones similares a las humanas, que mejoren la interacción humano-máquina.

1.1.1. Identificación de Problemas

En base a los avances en la modelación de las áreas corticales y subcorticales relacionadas con la toma de decisiones y aprendizaje, considerando la influencia de emociones y su potencial en su aplicación en plataformas robóticas, se plantea abordar los siguientes problemas:

1. ¿Cómo integrar efectos asociados a cambios en los niveles de DA tónica, neuromodulador relacionado a estados emocionales, en las interacciones de múltiples lazos cortico-ganglios basales?

2. ¿Cómo integrar un mecanismo de selección de acciones, que contemple aprendizaje e influencia de estados emocionales, en una plataforma robótica?

1.2. Trabajo previo

Diversos modelos que buscan definir el funcionamiento de estructuras cerebrales relacionadas con la toma de decisiones se han puesto a prueba en plataformas robóticas, lo que provee un medio para comprobar hipótesis teóricas y que puede conducir al desarrollo de mejores sistemas aplicados en agentes autónomos [28].

El modelo definido en [18] y [19] se ha implementado en tareas simples de supervivencia. Estas implementaciones, [8, 15–17, 31, 42], que no consideran aprendizaje, han demostrado que este tipo de modelos, comparado con un modelo clásico de *winner-takes-all* y reglas simples tipo *if-then-else*, presenta mejoras en cuanto a la resolución (evasión de indecisión) y persistencia del comportamiento, una de las propiedades fundamentales de mecanismos de selección de acciones [55]. Además, se ha extendido el uso de este modelo configurando sus parámetros basándose en técnicas de aprendizaje de máquinas, como algoritmos genéticos y redes neuronales, minimizando el establecimiento a priori de parámetros dependientes de la tarea a realizar por el agente [34, 56]. Sobre este modelo se ha integrado aprendizaje basado en la influencia de dopamina, con respecto a la relación entre acción y respuesta, no así sobre el valor de la respuesta para el agente, cuyo valor fue configurado a priori [5].

De forma similar, en [26] se ha integrado el aprendizaje basado en dopamina sobre conexiones cortico-corticales para relacionar acción y respuesta, utilizando un simple relé inhibitorio como modelo de ganglios basales.

En [27] se prueba el uso de neuromoduladores para controlar una plataforma robótica basada en el comportamiento. Este trabajo demuestra la capacidad de replicar el compor-

tamiento de un roedor basado en la influencia de neuromoduladores. Este modelo minimiza la integración de unidades corticales y subcorticales a la generación de neuromoduladores y su influencia en una unidad actuadora, y no considera aprendizaje.

En [52] se utiliza un modelo de ganglios basales y se controla el aprendizaje del valor de las acciones basado en dopamina, mas no considera la integración de emociones, caracteristica potencialmente beneficiosa para controladores de plataformas robóticas [32].

1.3. Contribuciones del trabajo

Las principales contribuciones de este trabajo de tesis son:

1. Se propone un modelo de las interacciones cortico-ganglios basales, conciliando diversas características presentes en la literatura. El modelo propuesto considera la presencia de múltiples niveles paralelos de toma de decisiones y aprendizaje del valor de las opciones basado en resultados, y efectos de los niveles de dopamina tónica traducidos en variaciones del comportamiento con respecto a la razón entre exploración y explotación.

2. Se define la estructura de un controlador para una plataforma robótica que integra como mecanismo de selección de acciones el modelo propuesto de lazos cortico-ganglios basales.

3. Se evalúa la integración de dopamina tónica, y por consecuencia características emocionales específicamente asociadas con trastornos de ansiedad, sobre el comportamiento de una plataforma robótica en la resolución de una tarea de supervivencia.

1.4. Estructura de la Tesis

Esta tesis se organiza como sigue:

- **Capítulo 2: Los ganglios basales.** Este capítulo describe y analiza el funcionamiento de los ganglios basales. Conjunto a una descripción de las estructuras que componen este grupo de núcleos neuronales, se describen de forma introductoria las rutas de conexiones neuronales que establecen el funcionamiento del mecanismo de toma de decisiones asociado a los ganglios basales.

- **Capítulo 3: Modelo computacional de los lazos cortico-ganglios basales.** Este capítulo describe el modelo propuesto que integra influencias del nivel de dopamina tónica sobre neuronas perteneciente a los ganglios basales. Específicamente, se consideran efectos sobre el estriado, estructura subcortical que actúa como entrada a los ganglios basales, modificando: las conexiones sinápticas entre éste y sectores de la corteza, la función de transferencia , Para evaluar el modelo, se presentan análisis en función del desempeño al realizar una tarea de decisión entre dos opciones simultáneas, con respecto a indicadores del nivel de aprendizaje y de los niveles de incertidumbre asociados al proceso de selección, caracterizando los efectos en términos de la razón de exploración-explotación.

- **Capítulo 4: Controlador bio-inspirado: Implementación del modelo propuesto como mecanismo de toma de decisiones.** Este capítulo define la estructura de un controlador de una plataforma robótica, cuyo mecanismo de toma de decisiones corresponde al modelo de lazos cortico-ganglios basales definido en el capítulo 3. El controlador es implementado para resolver una tarea de supervivencia básica, donde el agente autónomo aprende en función de sus requerimientos actuales qué acciones le recompensan, con el fín de mantener positivos niveles internos de vida. Para ello, la descripción de la tarea de supervivencia considera condiciones de recompensa tal que permiten que el robot sea capaz de aprender y seguir una estrategia que mantiene su funcionamiento en el tiempo. Sobre esta implementación se realizan análisis de los resultados en función de cómo se ve afectado el comportamiento del robot con respecto a los niveles de DA tónica, evaluando su capacidad de sobrevivir en el tiempo, las proporciones de las selecciones de cada alternativa disponible – cuyos cambios demuestran modulaciones sobre la estrategia de resolución del problema –, y de los niveles de exploración del ambiente.

- **Capítulo 5: Conclusiones.** Este capítulo relaciona los resultados obtenidos en los diferentes capítulos del documento, presenta un resumen del trabajo realizado, y propone guías para trabajo futuro.

- **Apéndice:** Este capítulo contiene los valores de los parámetros utilizados en las simulaciones desarrolladas en este trabajo.

Capítulo 2

GANGLIOS BASALES

Los ganglios basales son un conjunto de núcleos subcorticales ubicados en el centro del cerebro. Esta ubicación les permite estar ampliamente conectados con distintas zonas de los lóbulos frontales, formando lazos paralelos que procesan información según las zonas de la corteza a la que estén conectados. De esta forma, dependiendo de a qué zona esté conectado cada lazo, la decisión puede ser, por ejemplo, sobre el objetivo de un próximo movimiento sacádico, el objetivo de un movimiento de alcance, o qué información será guardada en memoria a corto plazo [16].

2.1. Anatomía

Las estructuras que componen los ganglios basales, presentadas en la figura 2.1, son el estriado, la sustancia negra pars compacta y pars reticulata, los globos pálidos interno y externo, y los núcleos subtalámicos.

Estriado (STR)

El estriado corresponde a la principal entrada de información hacia los ganglios basales. Presenta conexiones provenientes desde casi todas las areas de la corteza, excepto las zonas visual primaria y auditiva primaria [43]. En humanos, el estriado está compuesto principalmente por Neuronas Espinosas Medias (MSNs).

Las MSNs reciben conexiones desde distintas zonas de la corteza topográficamente distribuidas –de y hacia zonas específicas y separables–, manteniedo esta organización con respectivas conexiones de salida. Cada neurona recibe conexiones desde miles de neuronas pertenecientes a la corteza, por lo que integran la información transmitida por estas poblaciones de entrada. Estas neuronas (las MSNs) son además receptoras de dopamina, principalmente D1 y D2, neurotransmisores que se originan en la sustancia negra pars compacta y que modulan las entradas corticales. A pesar de la gran cantidad de conexiones de entrada, usualmente se encuentran en silencio, i.e., con un nivel de actividad aproximable a cero. Para ser activadas, las MSNs requieren que simultáneamente un gran número de estas conexiones se encuentren activas. Las MSNs son neuronas GABAérgicas, por lo que inhiben a las neuronas a las que se proyectan, pertenecientes al globo pálido y la sustancia negra pars reticulata.

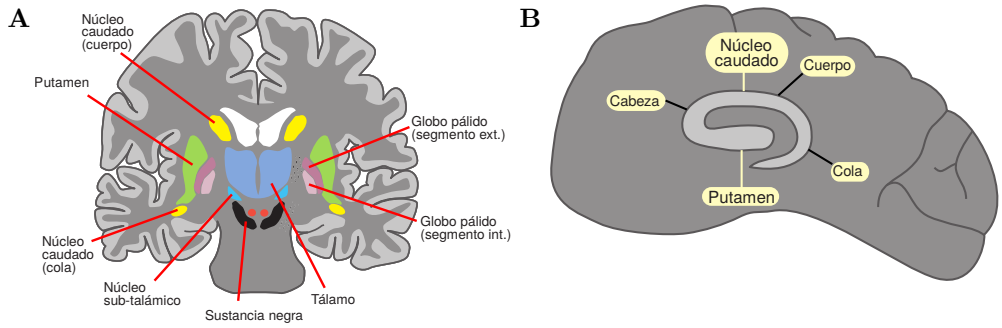

Figura 2.1: **Posición de los ganglios basales en el cerebro humano. A** Plano coronal. **B** Plano sagital. Figura adaptada de [51].

Sustancia negra pars compacta (SNc)

Conjunto de neuronas dopaminérgicas que afectan al estriado, modulando la activación de las MSNs.

Globo pálido interno (GPi) y sustancia negra pas reticulata (SNr): conjunto GPi/SNr

El globo pálido interno es usualmente integrado junto a la sustancia negra pars reticulata en un único conjunto. Este conjunto corresponde al núcleo principal de salida de los ganglios basales, compuesto por neuronas GABAérgicas (inhibitorias) tónicamente activas. Sus proyecciones se conectan de vuelta a la corteza a través del tálamo, todas conexiones topográficamente organizadas. Al ser neuronas inhibitorias la selección en los lazos cortico-ganglios basales se produce por desinhibición de neuronas del tálamo.

Globo pálido externo (GPe)

Núcleo compuesto por neuronas inhibitorias tónicamente activas, por lo que actúa como un negador intermedio de conexiones topográficamente organizadas desde el STR, con proyecciones hacia el núcleo subtalámico STN y el conjunto GPi/SNr.

Núcleo subtalámico (STN)

Neuronas excitatorias cuyas conexiones de entrada, desde el GPe y la corteza, presentan una organización topográfica. Sin embargo, las proyecciones de salida de las neuronas del STN son principalmente difusas, con conexiones hacia distintas poblaciones de neuronas del conjunto GPi/SNr.

2.1.1. Rutas de conectividad a través de los ganglios basales

En general, se describen tres rutas de conectividad en los ganglios basales [46], presentadas en la figura 2.2: una directa, una indirecta y una hiperdirecta.

- Ruta directa: considera las conexiones que conectan la corteza con el conjunto GPi/SNr a través del STR, inhibiendo la actividad de poblaciones del conjunto de forma topográficamente organizada. Define un lazo de realimentación positiva que promueve

la selección, para cada opción de la toma de decisiones en proceso, ya que al inhibir el conjunto GPi/SNr, aumenta la actividad en el tálamo y la corteza.

- Ruta indirecta: considera un lazo de realimentación negativa a través de dos sub-rutas, una corta que conecta el STR con el GPI/SNr a través del GPe, y otra larga que además considera una conexión entre el GPe y GPi/SNr a través del STN. Estas sub-rutas, en conjunto, tiene un efecto antagonista al efecto producido por la ruta directa, frenando la selección.

- Ruta hiper-directa: ruta que define un lazo de realimentación negativa con conexiones difusas sobre el GPi/SNr. A través de esta ruta, actividad en la corteza afecta de forma general a las poblaciones del GPi/SNr por medio de conexiones con el STN. Por lo tanto, para un lazo cortico-ganglios basales específico, actividad en neuronas del STN incrementan por igual la actividad de neuronas del conjunto GPi/SNr para todas las opciones, disminuyendo los niveles de actividad de todas las poblaciones del tálamo (asociadas al lazo específico).

2.2. Mecanismo de selección

El proceso de selección producido en los lazos cortico-ganglios basales es producto de una competición entre las rutas de conectividad definidas en la sección 2.1.1. La realimentación positiva de la ruta directa es fundamental para alcanzar un nivel de actividad cortical tal que se produzca una selección. Aparte de esta ruta, en la literatura se encuentran modelos que no consideran otras rutas competidoras, además de modelos que consideran competiciones con la ruta indirecta (corta, larga o ambas), con la ruta hyper-directa, o ambas [46].

Si bien la ruta indirecta aporta en el proceso de selección, diversos autores la consideran secundaria [29, 49]. El mecanismo propuesto por [29] considera las conexiones asociadas a las rutas directa e hiperdirecta, con el objetivo de describir síntomas del Parkinson observados en patrones de activación neuronal en los ganglios basales. La ruta directa presenta realimentación positiva y focalizada, mientras que la hiperdirecta presenta realimentación negativa y dispersa (ver figura 2.3). La competición entre ambas rutas permite realizar una selección entre varias opciones, similar al tipo *winner-takes-all*. A medida que aumenta la actividad cortical asociada a una opción, aumentando su posibilidad de ser seleccionada, inhibe con mayor intensidad al resto terminando en una única opción con destacada actividad cortical.

2.2.1. Extensión a múltiples selecciones

Mediante la integración de dos instancia del lazo propuesto por [29], [20] define un sistema de dos lazos paralelos simétricos que permiten realizar selecciones en dos niveles distintos, describiendo interacciones entre cada nivel para que actúen de forma coherente (ver figura 2.4). Estas interacciones se producen por la consideración de poblaciones asociativas en la corteza y el estriado: mientras que la corteza asociativa informa sobre la relación entre las opciones presentes de cada lazo (e.g., qué forma se encuentra dónde), el estriado asociativo

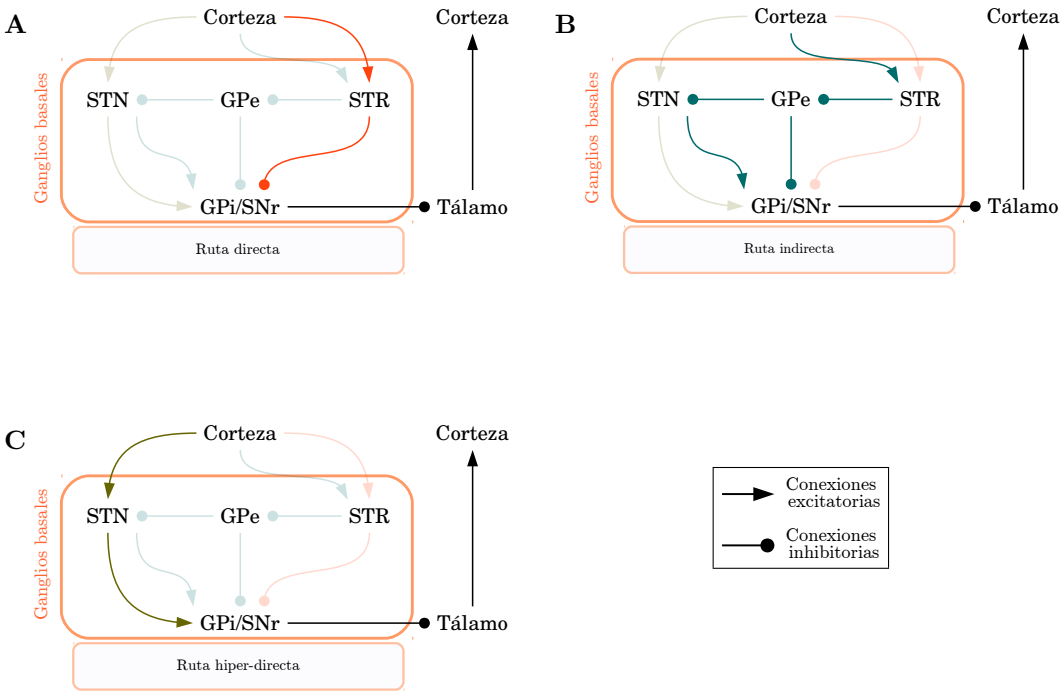

Figura 2.2: **Rutas de conectividad en los ganglios basales. A** Ruta directa: lazo de realimentación positiva en el cual aumentos en el nivel de actividad neuronal en la corteza, a través de conexiones excitatorias hacia el estriado (STR), produce inhibición en el conjunto de salida GPi/SNr. Ya que las conexiones desde el conjunto GPi/SNr inhiben el tálamo, al inhibir el GPi/SNr se favorece la actividad en el tálamo. **B** Ruta indirecta: lazo de realimentación negativo que frena los aumentos de actividad en el tálamo. Está compuesta por dos subrutas, una corta y una larga, las cuales conectan el estriado con el GPi/SNr a través del globo pálido externo (GPe). Esta conexión es directa (ruta *corta*) o a través del núcleo subtalámico (STN) (ruta *larga*). **C** Ruta hiper-directa: lazo de realimentación negativo, más rápido que los lazos anteriores. Conecta la corteza con el GPi/SNr a través del núcleo subtalámico, con conexiones difusas tipo *one-to-all*.

es responsable de la comunicación cruzada entre los lazos, permitiendo la influencia de uno sobre el otro.

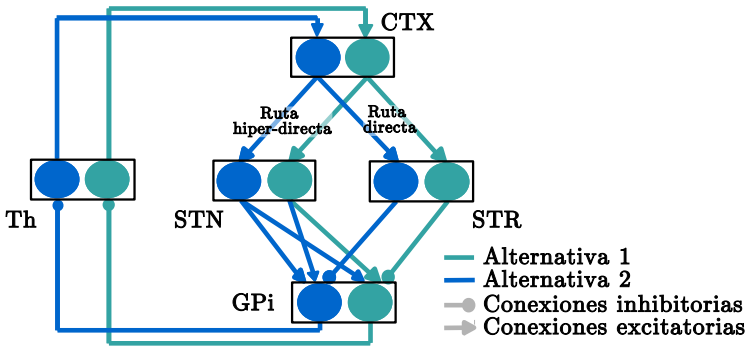

Figura 2.3: **Mecanismo de selección de acciones propuesto por Leblois et al.** El modelo incluye las rutas directa (CTX, STR, GPi y Th) e hiper-directa (CTX, STN, GPi y Th), considerando dos circuitos competidores (dos alternativas). La interacción entre estos circuitos se produce en las conexiones subtálamo pálidas, las cuales son difusas. El resto de las conexiones del modelo presentan una organización somatotópica – conexiones punto a punto –. Figura adaptada de [29].

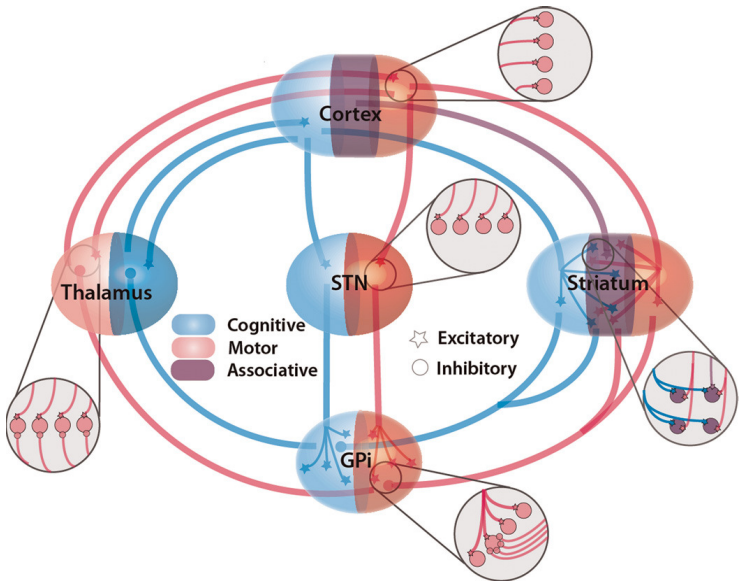

Figura 2.4: **Modelo de los lazos cortico-ganglios basales propuesto por Guthrie et al.** Modelo que considera dos lazos cortico-ganglios basales paralelos, uno asociado a una decisión cognitiva (en azul) y otro asociado a una decisión motora (en rojo). Cada lazo es una instancia del mecanismo de selección de acciones propuesto por de [29] (ver figura 2.3). Mediante la integración de poblaciones neuronales asociativas, en la CTX y en el STR, existe una comunicación entre los lazos que permite realizar decisiones coherentes entre ambos dominios, cognitivo y motor. Este modelo considera aprendizaje sobre los pesos de las conexiones cortico-estriadas del lazo cognitivo basado en las variaciones fásicas de dopamina, las cuales contienen información de error entre respuesta esperada y obtenida. Figura de [20].

Capítulo 3

MODELO COMPUTACIONAL DE LOS LAZO CORTICO-GANGLIOS BASALES

3.1. Introducción

El presente desarrollo corresponde a una extensión del modelo propuesto por Guthrie et al. [20], en el cual se detallan las interacciones cortico-ganglio basales como un sistema de múltiples lazos paralelos. Los modelos de múltiples lazos paralelos proponen la existencia de varios lazos segregados a través de los ganglios basales, que se conectan con distintas áreas corticales, cada una asociada a un dominio de selección distinto [3]. De esta forma, se pueden discriminar lazos motores, oculomotores, asociados a la corteza prefrontal y a las estructuras del sistema límbico [2]. El trabajo expuesto por [20] define las interacciones entre estos lazos que permiten la selección de acciones en más de un nivel o dominio. Con ello, su modelo computacional permite simular actividad cerebral registrada en ganglios basales de primates [41] que sugiere la existencia de dos procesos separables, uno de selección visual o cognitiva (entre formas) y otro visual-espacial o motora (entre direcciones de movimiento).

3.1.1. Innovación del modelo propuesto

Del modelo de Guthrie et al. [20], se consideran las ecuaciones que describen las dinámicas del sistema de lazos múltiples, además de considerar aprendizaje dopaminérgio en conexiones cortico-estriadas. Sobre estas características, se integran influencias de los niveles de dopamina tónica en las neuronas del estriado.

Por lo tanto, el modelo propuesto considera:

- A partir del modelo de Guthrie et al.:
  - La definición de dos lazos paralelos. La descripción de los lazos se realiza en detalle en la sección 3.2.
  - Aprendizaje dopaminérgico basado en recompensas, aplicado sobre las conexiones cortico-estriadas pertenecientes a uno de los lazos. El proceso de aprendizaje es descrito en la sección 3.2.1.

- Integración de efectos asociados a los niveles de dopamina tónica en las neuronas del estriado, relacionados con modulación del comportamiento con respecto a la razon de exploración-explotación en los procesos de selección [22], descritos en la sección 3.2.2:
  - Modulación de los niveles de actividad neuronal provenientes de las conexiones con la corteza.
  - Modificación de los parámetros que definen la función de transferencia asociada a las neuronas del estriado.
  - Regulación de la razón señal-ruido.

En la sección 3.3, el modelo propuesto es puesto a prueba en una tarea de selección forzada con dos opciones, misma tarea utilizada para evaluar el modelo de [20]. La tarea utilizada se aplica en formato de realizaciones individuales y en formato de realizaciones consecutivas considerando aprendizaje. En base a los resultados obtenidos, en la sección 3.4 se presenta una discusión sobre los efectos observados para distintos niveles de DA tónica. La sección 3.5 presenta conclusiones sobre el modelo, la implementación y los resultados obtenidos.

3.2. Descripción del modelo computacional

En la figura 3.1 se presenta el modelo con todas las estructuras, corticales y subcorticales, de ambos lazos implementados. El dominio de la información sobre el cuál cada lazo realiza una selección es definido por el tipo de información de las entradas, provenientes de la corteza. En la presente implementación, como en [20], se considera un lazo realizando una selección sobre un dominio cognitivo y otro lazo sobre un dominio motor. El lazo cognitivo realiza una selección entre reconocimientos cognitivos de las opciones disponibles, de aquí en adelante llamadas *alternativas*, basando la selección entre los beneficios esperados de cada opción. De forma similar y en paralelo, el lazo motor realiza una selección a nivel de *acciones*, entre los movimientos físicos asociados a la opción seleccionada. Un ejemplo de esta representación paralela es la selección entre dos figuras en una pantalla: mientras que la selección a nivel cognitivo puede ser basada en las formas de las figuras, la selección motora puede ser apuntar en la dirección de alguna de éstas, indicando qué figura se ha seleccionado. Para que este proceso de selección se desarrolle de forma correcta, la decisión a nivel cognitivo debe guiar la decisión a nivel motor.

En cada estructura, cortical y sub-cortical, se considera un grupo de neuronas co-activadas asociadas a una opción específica, denominado ensamble. En el caso de las estructuras asociativas, se considera un grupo por cada par *alternativa-acción*. Cada uno de estos grupos es simulado mediante un modelo neuronal tipo *neural rate* [59], representado por una única neurona. La dinámica de las actividades sinápticas aferentes son descritas por:

$$\tau \frac{dm_i(t)}{dt} = -m_i(t) + \mu_i(t),$$
$$\mu_i(t) = S(I_i^T(t) - T_i),$$

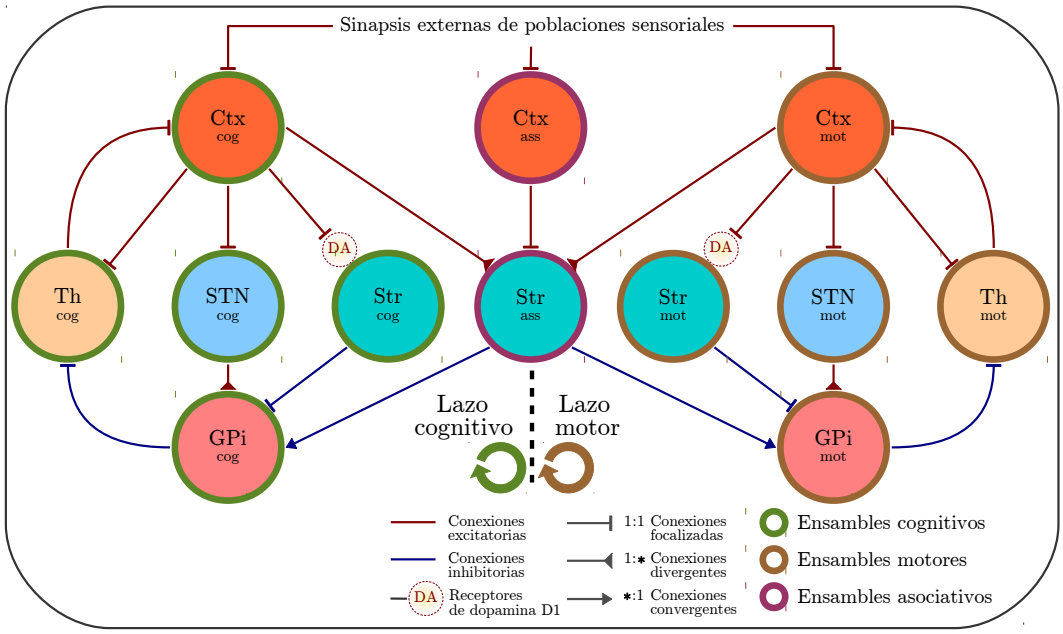

Figura 3.1: **Lazos cortico-ganglio basales con integración de influencia de DA tónica.** Sobre el modelo propuesto por Guthrie et al. (ver figura 2.4), se incluye la influencia de DA D1 tónica en neuronas estriadas, estructura con un predominante número de neuronas receptoras de DA. El modelo considera dos lazos simétricos, que interactúan entre ellos a través de estructuras asociativas, pertenecientes a la CTX y el STR. Las conexiones motoras y cognitivas cortico estriadas ($CTX \rightarrow STR$), cortico subtalámicas ($CTX \rightarrow STN$), estriado pálidas ($STR \rightarrow GPi$), pálido talámicas ($GPi \rightarrow Th$) y tálamo corticales ($Th \leftrightarrow CTX$) presentan una estructura topográfica – punto a punto –. Las conexiones subtalamo pálidas ($STN \rightarrow GPi$) son divergentes, parte de la ruta hiper-directa, produciendo una realimentación negativa difusa: las unidades del STN excitan a todas las unidades del GPi, de su lazo respectivo, que a su vez inhibe al Th. Las estructuras asociativas, que contienen ensambles para cada par cognitivo-motor posible, presentan: conexiones cortico estriadas asociativas de tipo topográficas; conexiones cortico estriadas cognitivas y motoras de tipo divergente, donde cada ensamble cortical cognitivo afecta a todos los ensambles estriados motores asociativos que conectan a esa alternativa (pares *alternativa* específica - todas las *acciones*) y viceversa; y conexiones estriado pálidas ($STR \rightarrow GPi$) convergentes, donde cada par asociado a una alternativa cognitiva (motor) afectará al respectivo ensamble cognitivo (motor). Los efectos considerados asociados a la DA tónica en neuronas estriadas, descritos en la sección 3.2.2, son: modulación sobre la fuerza de las conexiones cortico estriadas, modificación de los umbrales de activación de las neuronas estriadas, y variaciones sobre la relación señal-ruido.

donde i corresponde al ensamble, con $i \in \{CTX_e, STR_e, STN_e, GPi_e, Th_e\}$ para los niveles $e \in \{ass, cog, mot\}$ para las estructuras pertenecientes al STR y la CTX, y los niveles $e \in \{cog, mot\}$ para el resto de las estructuras: STN, GPi y Th. τ es una constante de tiempo de decaimiento asociada a las entradas sinápticas. $m_i(t)$ es la actividad sináptica de salida, correspondiente al paso de la actividad neuronal instantánea $\mu_i(t)$ por un filtro pasabajos. $S(\cdot)$ es la función de transferencia no lineal del ensamble, $I_i^T(t)$ es la entrada sináptica total y T_i es un umbral tal que la actividad de entrada debe ser mayor a éste para generar actividad sináptica en el ensamble. Por simplicidad, excepto para las unidades pertenecientes al STR ya que la función de transferencia que simula su comportamiento es

distinta (ver ecuación (3.2.1)),

$$S(x) = \begin{cases} x & \text{si } x \geq 0, \\ 0 & \text{otros casos} \end{cases}$$

Además, los umbrales T_i son iguales cada ensamble perteneciente a i, es decir, las distintas opciones representadas en cada estructura y nivel (cognitivo, asociativo o motor) presentan el mismo umbral de activación.

Para considerar fluctuaciones estocásticas en las conexiones neuronales, se inserta ruido en todas las entradas sinápticas. Entonces, las entradas sinápticas totales para un ensamble i son

$$I_i^T(t) = I_i^S(t) + \eta_i(t),$$

donde $\eta_i(t) \sim \mathcal{N}(0, \lambda_i I_i^S)$ corresponde a realizaciones pseudo-aleatorias de un proceso Gaussiano con desviación estándar λ_i veces proporcional a la actividad de entrada I_i^S.

Los ensambles pertenecientes a la corteza reciben además actividades de entrada provenientes de sectores sensoriales externos al modelo: I_i^{Ext}. Por lo tanto, para $i \in \{Ctx_{cog}, Ctx_{mot}\}$, la actividad total de entrada resulta

$$I_i^T(t) = I_i^S(t) + I_i^{Ext}(t) + \eta_i(t),$$

con

$$I_i^{Ext}(t) = \begin{cases} I_{EXT}(t) + \eta_{i,EXT} & \text{si opción está presente,} \\ 0 & \text{en caso contrario,} \end{cases}$$

donde $I_{EXT}(t)$ es una función que determina la forma de onda de la actividad proveniente de la CTX sensorial (función cuadrada en [20]), y $\eta_{i,EXT} \sim \mathcal{N}(0, \lambda_{Ctx} \cdot I_{EXT})$ corresponde a una realización de ruido Gaussiano, computado una única vez al ser presentada la opción asociada. Para los ensambles corticales cognitivos, motores y associativos, $I_i^{Ext}(t)$ representa la disponibilidad de cada *alternativa*, cada *acción* y cada asociación *alternativa-acción*, respectivamente.

La actividad sceptical para los ensambles pertenecientes al estriado se simula mediante la aplicación de una función de transferencia de tipo sigmoide, según

$$S_{Str}(I_i^S(t)) = V_{min} + \left(\frac{V_{max} - V_{min}}{1 + e^{\frac{V_h - I_i^T(t)}{V_c}}} \right), \tag{3.2.1}$$

donde $I_i^S(t)$ es la actividad sináptica proveniente de los ensambles corticales, V_{min} es el valor mínimo de activación, V_{max} es la activación máxima, V_h es la actividad de entrada para la cual se produce la mitad de la activación máxima, y V_c es la pendiente. Esta descripción considera que las principales neuronas aferentes del estriado, las Neuronas Espinosas Medias (MSNs):

i) Actividad de entrada orquestada para ser activadas sobre su nivel en resposo [58].

II) Mientras se encuentran en reposo presentan una baja actividad sináptica (sinapsis silensiosas) [45].

III) Su potencial de membrana se comporta de acuerdo a una función de tipo sigmoide sobre el total de corrientes de entrada [37]

El comportamiento resultante al aplicar (3.2.1) es de nula o poca actividad sináptica para pequeñas actividades de entrada, rápidamente aumentando hasta una activación igual a V_{max} para entradas con mayor intensidad.

Las entradas sinápticas desde ensambles pre-sinápticos j hacia un ensamble post-sináptico i, I_i^S, se descompone como

$$I_i^S(t) = \sum_j G_i^j m_j(t), \tag{3.2.2}$$

donde G_i^j es la ganancia de la conexión sináptica entre las poblaciones j e i.

Se determina que el modelo realiza una selección una vez que un ensamble perteneciente a las poblaciones de la CTX motora alcanza un nivel de actividad sináptica de al menos 40 $[sp/s]$ mayor a cualquier otra *acción* disponible.

3.2.1. Aprendizaje dopaminérgico en conexiones cortico-estriadas

El modelo propuesto por Guthrie et al. considera aprendizaje basado en variaciones fásicas del nivel de dopamina en los pesos correspondientes a las conexiones cortico-estriadas del lazo cognitivo. La aplicación de aprendizaje sólo en el lazo cognitivo es una implementación razonable considerando que, dado el dominidio de decisión, la selección cognitiva debe guiar a la decisión motora tal que sean coherentes. Incluso considerando que fuese un posición espacial el dominio asociado al proceso de selección, la internalización de esta cualidad (posición de los objetos, por ejemplo) correspondería a una proceso de caracterización cognitiva.

Para describir este proceso de aprendizaje, se extiende la descripción de la ecuación (3.2.2) para las conexiones cortico-estriadas cognitivas de forma tal que los pesos de las conexiones sinápticas w_i^j, entre las poblaciones j e i, sean observables. Esto, según

$$I_i^S(t) = \sum_j w_i^j \cdot G_i^j m_j(t),$$

El aprendizaje se lleva a cabo en función de la presencia, o nó, de una recompensa $R \in \{0, 1\}$ una vez completada la ejecución de una *acción* previamente seleccionada. Esta evaluación corrige valores internos V_k asociados a cada *alternativa* k, los cuales simulan memoria de largo plazo. Cada valor V_k representa un nivel de beneficio de la respuesta obtenida por haber seleccionado la *alternativa* k, después de haber ejecutado la respectiva *acción* asociada.

El proceso de aprendizaje se realiza aplicando un algoritmo *actor-critic*, parte de los algoritmos de aprendizaje desarrollados en el área de Aprendizaje Reforzado [53]. Son varios los autores que han asociado técnicas de Aprendizaje Reforzado, propuestas desde puntos de vista asociados a la ingeniería y la Inteligencia Artificial, con modelos de actividad neuronal en sistemas con aprendizaje dopaminérgico. Entre ellos, se encuentran modelos que describen

este aprendizaje con algoritmos de *temporal-difference* [21, 38, 47], y modelos que integran percepción de valor esperado con algoritmos *actor-critic*.

La implementación del algoritmo de aprendizaje sigue la siguiente descripción:

$$PE = R - V_k,$$
$$w_{Str_{cog}}^{Ctx_{cog}} \leftarrow PE \cdot \alpha_a \cdot m_{Str_{cog}}(t) + w_{Str_{cog}}^{Ctx_{cog}}, \tag{3.2.3}$$
$$V_k \leftarrow PE \cdot \alpha_c + V_k, \tag{3.2.4}$$

donde PE es la predicción del error: diferencia entre el valor recibido como recompensa R y el valor esperado V_k de la *alternativa* previamente seleccionada. Basado en el valor de PE, se corrigen los pesos de la conexión cortico-estriada cognitiva $w_{Str_{cog}}^{Ctx_{cog}}$, asociada a k, considerando una tasa de aprendizaje para el actor α_a modulada por la actividad sináptica del respectivo ensamble del estriado cognitivo $m_{Str_{cog}}$. El valor esperado V_k es corregido aplicando un factor α_c, la taza de aprendizaje del crítico.

3.2.2. Integración de la influencia de DA D1 tónica sobre los lazos cortico-ganglios basales

Para integrar efectos del nivel de DA tónica en el modelo, se considera que las neuronas estriadas de la ruta directa son receptoras principalmente excitatorias para DA tipo D1 [14], cuyo efecto es caracterizable como factor multiplicativo [19]. Basado en ello, se modifica el cálculo de la actividad sináptica de entrada total $I_i^T(t)$ descrita en la ecuación (3.2.1), para el lazo cognitivo y motor en las poblaciones neuronales del estriado, según:

$$I_i^T(t) = (1 + DA)\ I_i^S(t) + \eta_i(t),$$

donde DA es el nivel de DA tónica tipo D1. Siguiendo una implementación similar, en [22] se determina que efectos asociados a los niveles de DA tipo D1 son suficientes para modular comportamientos de selección en lazos cortico-ganglios basales en términos de la razón entre exploración y explotación.

En estudios *in vitro* se han relacionado incrementos en el umbral de activación de neuronas del estriado con incrementos en el nivel de DA tónica [6,39]. Esto se integra describiendo el umbral V_h de la función de transferencia de las neuronas estriadas cognitivas y motoras (ecuación (3.2.1)) como linealmente dependiente del nivel DA, según:

$$V_h = V_{h_{DA}}\ DA + \check{V}_h, \tag{3.2.5}$$

donde $V_{h_{DA}}$ es el factor proporcional con respecto a DA y $\check{V}_h$ es el mínimo umbral, presente para $DA = 0$.

Con el fín de mantener el rango de actividad sináptica de entrada, en conjunto a la modificación del umbral V_h, la pendiente V_c de la función de transferencia del estriado cognitivo y motor se define en función del nivel de DA tónica, implementando la relación lineal

$$V_c = V_{c_{DA}}\ (1 + DA), \tag{3.2.6}$$

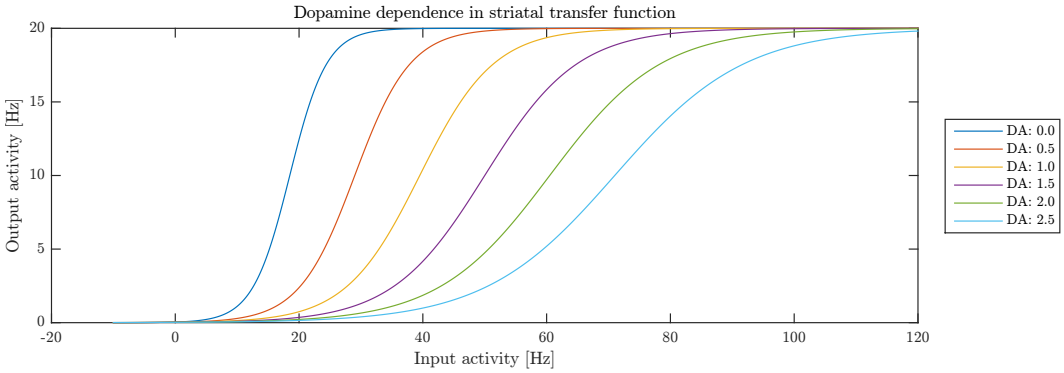

Figura 3.2: **Efectos de la DA tónica en la función de transferencia del estriado, ecuación** (3.2.1)**.** Efectos de los niveles de DA tónica en las poblaciones cognitivas y motoras del estriado, considerando cambios en el umbral de activación y en la pendiente, descritos en las ecuaciones (3.2.5) y (3.2.6), respectivamente.

donde $V_{c_{DA}}$ es un factor proporcional. Al ser la pendiente ajustada por $(1 + DA)$, mismo factor de escala sobre la actividad sináptica de entrada en la ecuación (3.2.5), la influencia de V_c en el nivel de actividad de entrada $I_i^S(t)$ (previa a la consideración de DA) es independiente del nivel de DA tónica.

Además de los efectos considerados en las intensidades de entrada y en la función de transferencia de las neuronas estriadas, incrementos en el nivel de DA tónica genera incrementos en la relación señal-ruido (SNR) [36, 39]. Por ello, se considera que el nivel de ruido de entrada para las poblaciones estriadas de los lazos cognitivo y motor como independiente del nivel de entrada sináptica actual. Adicionalmente, considerando que las neuronas aferentes del estriado requieren una fuerte correlación en la entrada para verse afectadas [12], al ruido aplicado en el estriado cognitivo y motor se le aplica un filtro pasabajos para insertar considera correlación temporal. Con esta correlación presente, el sentido del ruido se mantiene estable el tiempo suficiente para guiar la selección. Entonces, la dinámica que describe el ruido de entrada para el estriado cognitivo y motor es:

$$\tau_{\eta_{Str}} \frac{d\eta_{Str_e}(t)}{dt} = -\eta_{Str_e}(t) + n_{Str_e}(t), \tag{3.2.7}$$

para $e = \{cog, mot\}$, donde $\tau_{\eta_{Str}}$ es la constante de tiempo de decaimiento y $n_{Str_e}(t)$ corresponde a ruido Gaussiano con un valor de varianza constante: $n_i \sim \mathcal{N}(0, \lambda_{Str_{\substack{cog \\ mot}}})$.

3.3. Evaluación de los effectos de la DA tónica en el modelo propuesto

Con el objetivo de caracterizar los efectos producidos por variaciones en el nivel de DA tónica en el modelo de dos lazos cortico-ganglios basales detallado en la sección 3.2, se simula la ejecución de una tarea de aprendizaje de selección forzada con dos opciones, misma a la utilizada en [20]. El experimento consiste en realizar una selección entre dos figuras en una pantalla, donde la forma de las figuras presentadas tiene asociada una probabilidad de recompensa constante. Esta tarea permite evaluar los efectos dopaminérgicos integrados al sistema durante un proceso de aprendizaje a lo largo de un determinado número de

selecciones consecutivas.

3.3.1. Tarea de selección forzada con dos opciones

La tarea utilizada para caracterizar cómo se ve afectado el modelo propuesto en función de niveles simulados de DA tónica en el estriado, con el fín de determinar si existe una modulación de la razón de exploración-explotación en el proceso de selección, corresponde a una *tarea de selección forzada con dos opciones*. Específicamente, se simula una tarea correspondiente a realizaciones consecutivas de un *trial* donde se presentan dos figuras diferentes, de las cuales se debe seleccionar una dentro de un tiempo determinado, considerando aprendizaje entre cada *trial*.

Composición de cada trial

Cada *trial* parte con un tiempo de estabilización de 500 $[ms]$, donde el modelo alcanza un estado estacionario en cada ensamble, en términos de la actividad neuronal media. Al instante en que termina el periodo de estabilización se presentan dos figuras diferentes, elegidas de un conjunto de cuatro figuras distintas, de forma pseudo-aleatoria siguiendo una distribución uniforme. Las figuras son presentadas en dos posiciones diferentes, elegidas de forma equivalente de un total de cuatro posiciones posibles, siendo ambas selecciones (forma y posición) independientes. Durante los siguientes 2500 $[ms]$, puede realizarse una selección si un ensamble cortical motor presenta actividad sináptica superior a 40 $[sp/s]$ en comparación a cualquier otro ensamble de la misma estructura motora, considerando como selección realizada en el primer instante en que se cumple esta condición. Una vez transcurrido este periodo de selección se retiran las figuras, y cualquier selección posterior es desechada. El *trial* termina con un segundo periodo de estabilización de 500 $[ms]$. Este proceso se presenta de forma gráfica en la figura 3.3.

La función utilizada como $I_{EXT}(t)$ aplicada en la ecuación (3.2.1), correspondiente a la actividad sináptica proveniente de sectores de corteza sensoriales como entrada a los ensambles de CTX del modelo (ver figura 3.1), se compone por dos sigmoides con el fín de generar una señal cuadrada con transiciones suaves (ver figura 3.4). Matemáticamente, se describe como

$$I_i^{Ext}(t) = A\left(\frac{1.0}{1.0 + e^{\frac{0.725-t}{0.042}}} - \frac{1.0}{1.0 + e^{\frac{3.2-t}{0.084}}}\right), \tag{3.3.1}$$

donde A es la amplitud máxima.

Recompensas

Como se describe en la sección 3.2.1, el modelo considera una regla de aprendizaje dopaminérgico siguiendo un algoritmo tipo *actor-critic*, donde el sistema fortalece o debilita las conexiones cortico-estriadas asociadas a la *alternativa* seleccionada en función de recibir de recompensa luego de haber completado la ejecución de la *acción* correspondiente. Si se recibe una recompensa mejor que la esperada, las conexiones son fortalecidas. Al contrario, si se recibe una recompensa peor de la esperada las conexiones se debilitan.

Para cada figura de las *alternativas* posibles (ver figura 3.3), con $k \in \{1, 2, 3, 4\}$ índice de cada figura, se determinan probabilidades de recompensa $P(R|k)$, probabilidad de ser recompensado habiendo seleccionado la *alternativa* k, según:

$$P(R|k) = (4 - k)/3 \tag{3.3.2}$$

Por lo tanto, al seleccionar la figura 1 la probabilidad de recompensa $P(R|k = 1)$ es igual a 1, mientras que para $k = 2$ es igual a $2/3$, para $k = 3$ es $1/3$ y para $k = 4$ la probabilidad de recompensa es nula. Se espera entonces que el sistema, a lo largo del proceso de aprendizaje modifique su comportamiento aumentando la probabilidad de seleccionar la figura asociada al índice menor, dada la presentación de dos figuras distintas: seleccionar la figura 1 frente a cualquier otra, la figura 2 frente a las figuras 3 y 4, y la figura 3 frente a la figura 4.

Evaluación de desempeño

Para evaluar el desempeño en la realización de la tarea expuesta, de selección forzada considerando aprendizaje, se utiliza una medición definida como el *desempeño medio* $M_p(n_t)$, promedio normalizado de *trials* exitosos. Se define un *trial* como exitoso si es que se realiza una selección de la figura con mayor probabilidad de recompensa. En detalle, para $n_t \in [1, 240]$ el número del *trial* actual, el desempeño $p(n_t)$ y el desempeño medio $M_p(n_t)$ se definen como:

$$\begin{aligned} p(n_t) &= \begin{cases} 1 & \text{if } P(R|k_s) > P(R|k_r) \\ 0 & \text{otherwise,} \end{cases} \\ M_p(n_t) &= \frac{1}{n} \sum_{i=1}^{n_t} p(i), \end{aligned} \tag{3.3.3}$$

donde k_s es el índice de la figura seleccionada, y k_r es el índice de la figura no seleccionada.

3.3.2. Determinación de parámetros integrados al modelo

Una vez definido un rango aproximado de DA tónica a utilizar ($DA \in [0.0, 2.5]$), mediante la evaluación de *trials* individuales considerando ciertas condiciones a detallar, se determinan los parámetros asociados a la influencia de la DA tónica en el modelo:

- $V_{h_{DA}}$: El factor proporcional entre DA y el umbral de la función de transferencia del estriado descrito en la ecuación (3.2.5).
- $V_{c_{DA}}$: El factor asociado a la pendiente, descrito en la ecuación (3.2.6).
- $\lambda_{Str_{\substack{cog \\ mot}}}$: El nivel de varianza del ruido de entrada en el estriado $n_{Str_e}(t)$, descrito en la ecuación (3.2.7).

Las condiciones utilizadas se basan en un funcionamiento normal del modelo. Estas características se detallan a continuación, especificando los principales parámetros que se condicionan por cada requerimiento:

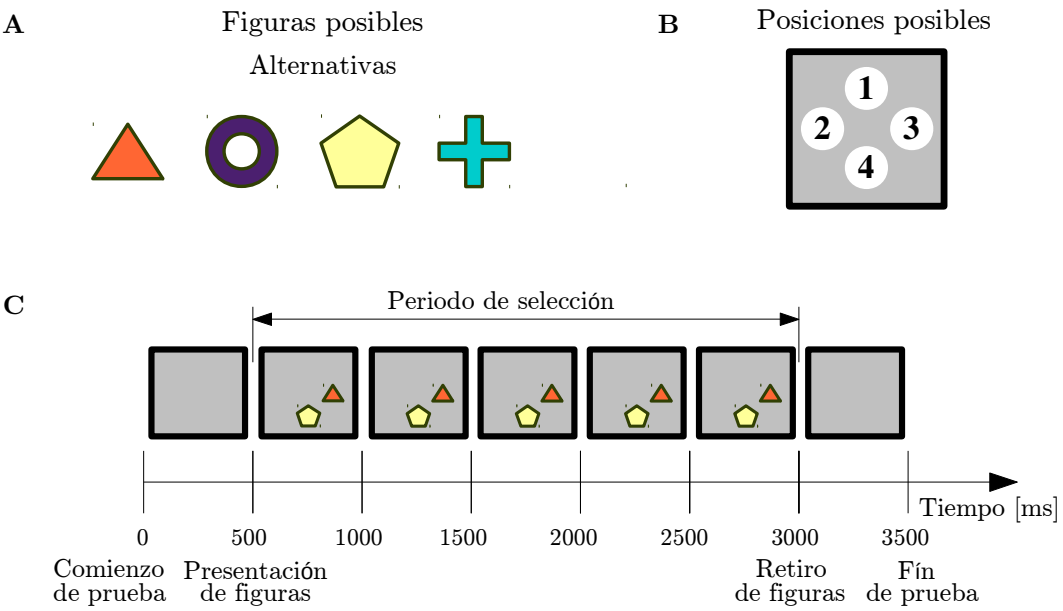

Figura 3.3: **Representación de un *trial* de la tarea de selección.** La tarea, descrita en forma detallada en la sección 3.3.1, está compuesta de realizaciones consecutivas de un *trial*. Cada *trial* consiste en la presentación de dos opciones durante un periodo de 2500 [*ms*], durante el cuál se debe indicar una figura seleccionada, admitiendo realizaciones sin selección alguna. **A** Cuatro formas diferentes que corresponden a las *alternativas* asociadas a las entradas del lazo cognitivo, siendo dos las presentadas en forma simultánea. **B** Cuatro posiciones diferentes posibles, sobre las cuales se presentan las *alternativas*, representando cuatro *acciones* disponibles para realizar la selección, asociadas a las entradas del lazo motor. **C** Ejemplo gráfico, a través del tiempo, de la ejecución de un *trial* (realización de una instancia de la tarea), donde dos figuras diferentes son presentadas (triángulo y pentágono) en dos posiciones diferentes (respectivamente: derecha, posición 3; y abajo, posición 4).

- Capacidad de selección: una vez presentadas las opciones disponibles, el nivel de actividad cortical debe ser suficiente como para activar el estriado y disparar el proceso de selección. Esto determina un nivel mínimo de la amplitud de $I_{EXT}(t)$, A en la ecuación (3.3.1), conjunto a un nivel máximo de umbral en el estriado tal que permita su activación con el nivel de actividad cortical presente.

- Quiebre espontáneo de simetría: se evita que se ejecuten selecciones sin tener opciones presentes, i.e., con $I_{EXT}(t) = 0$ para toda población cortical. Esto limita el nivel de ruido tal que sea insuficiente para que se dispare de forma espontánea una selección. Es importante tener en cuenta que dada la naturaleza tipo *winner-takes-all* del modelo, existe un nivel de diferencia de actividad cortical tal que se dispara el proceso que conlleva posteriormente a la selección, ya que el sistema presenta una realimentación positiva con ganancia mayor a 1 en la ruta directa.

- Persistencia en actividad cortical: una vez retiradas las opciones, la actividad cortical debe decaer hasta alcanzar nuevamente un estado estacionario con actividad media. Esto determina un nivel mínimo para $V_{h_{DA}}$ tal que la actividad cortical remanente no

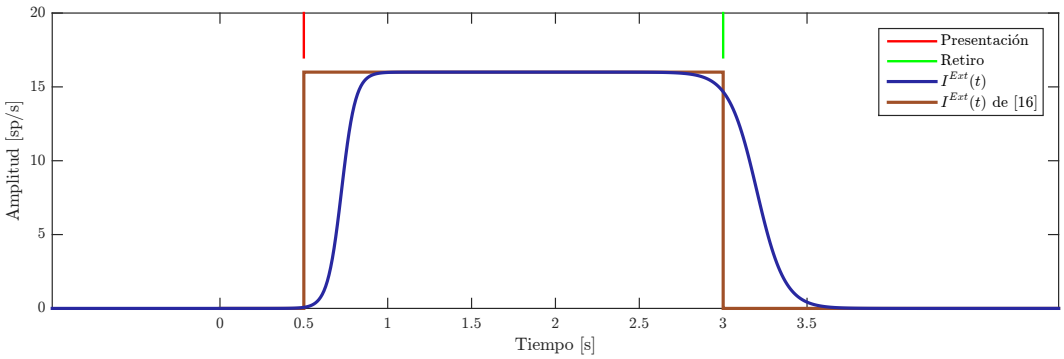

Figura 3.4: **Forma de onda de la actividad sináptica externa.** Forma de la actividad sináptica externa de entrada a los ensambles de corteza, $I^{Ext}(t)$, definida en (3.3.1). En [20] esta actividad es simulada como una señal cuadrada con estados binarios con cambios abruptos en los instantes de presentación y retiro de las figuras. En este trabajo se toman en consideración 300 $[ms]$ de procesamiento relacionado con procesos de percepción que suavizan la transición entre los niveles de actividad.

sea suficiente para mantener los niveles de actividad en un estado de selección.

- Velocidad de selección: el requerido para realizar una selección debe estar contenido en el tiempo establecido como posible (2500 $[ms]$ del periodo de selección, figura 3.3). Esto influye en el ajuste del factor asociado a la pendiente en el estriado, $V_{c_{DA}}$, ya que establece una relación entre entre variaciones de actividad cortical y variaciones de actividad resultante en el estriado.

Estas condiciones se deben cumplir para cualquier combinación de pesos cortico-estriados cognitivos afectos por el aprendizaje implementado. Por lo tanto, se realizan simulaciones considerando como estado inicial un estado previo a aprendizaje y un estado post-aprendizaje. Los niveles de pesos post-aprendizaje utilizados fueron adquiridos a través de simulaciones utilizando los parámetros del modelo original de Guthrie et al., sin considerar la integración de DA tónica.

Utilizando los parámetros determinados, en la figura 3.2 se presentan los efectos de la DA tónica, descritos en las ecuaciones (3.2.5) y (3.2.6), sobre la función de transferencia del estriado cognitivo y motor. En la figura 3.5 se presentan como ejemplos realizaciones de *trials* para cada nivel de DA considerado, para un estado del sistema previo a aprendizaje.

3.3.3. Resultados en la ejecución de la tarea de selección durante la fase de aprendizaje

Se realizan 300 simulaciones de la tarea de selección forzada para cada nivel de DA tónica, presentando en el tiempo el desempeño obtenido. Con esto, se espera determinar si el sistema es capaz de aprender, para cada nivel de DA, a seleccionar las figuras que tengan mayor probabilidad de generar una recompensa. La capacidad de aprendizaje se mide directamente por el desempeño medio, definido en la ecuación (3.3.3), valor que aumenta si el agente selecciona con mayor frecuencia la figura presentada con mayor probabilidad de

recompensa. Los niveles de desempeño a lo largo de simulaciones, con un total de 240 *trials*, son presentados en la figura 3.6.

Otra característica que se puede distinguir en las curvas de desempeño medio es la distancia entre el valor medio (0.5), alcanzable si la selección es aleatoria con igual probabilidad de seleccionar cada figura, y el valor alcanzado de desempeño. Esta distancia permite evaluar de forma preliminar si el proceso de selección presenta un comportamiento más inclinado hacia la explotación (con valores más lejanos al medio) o hacia la exploración (con valores más cercanos al medio). Es importante comentar que esta evaluación es simplemente preliminar y no concluyente, ya que el sistema puede estar seleccionando para ciertos pares de figuras de forma correcta y para otros de forma incorrecta, de forma consistente, resultando también en un valor de desempeño medio cercano a 0.5.

Los valores medios de los desempeños medios alcanzados, para cada nivel de DA tónica, son presentados en la tabla 3.1

Con el objetivo de asegurar un correcto aprendizaje con respecto a la percepción del valor de las figuras, se presenta en la figura 3.7 la evolución de V_k, valor interno asociado a memoria a largo plazo descrito en la ecuación (3.2.4). Para cada nivel de DA, los valores alcanzados de forma asintótica corresponden a los definidos en (3.3.2).

De forma similar, en la figura 3.8 se presenta la evolución de los pesos cortico-estriados del lazo cognitivo, a lo largo del proceso de aprendizaje. Es importante destacar que los pesos alcanzados dependen directamente de la tasa de aprendizaje α_a y de la evolución de los valores internos para cada figura V_k (ver ecuación (3.2.3)). Los pesos resultantes del proceso de aprendizaje, presentados en la tabla 3.2, son similares para cada nivel de DA utilizado.

3.3.4. Resultados en la ejecución de la tarea de selección habiendo finalizada la fase de aprendizaje

Simulando una etapa posterior al proceso de aprendizaje, se prueba el modelo en repeticiones de *trials* individuales, con el objetivo de adquirir estadísticas asociadas a la selección. Durante la realización de estos experimentos, se utilizan los valores medios de los pesos aprendidos (ver sección 3.3.3 para cada nivel de DA tónica evaluado, manteniendo los niveles de DA de forma coherente.

DA tónica	0.0	0.5	1.0	1.5	2.0	2.5
$M_p(n_t = 240)$	0.73372	0.78471	0.82111	0.84036	0.85635	0.86844

Tabla 3.1: **Desempeño medio final para cada nivel de DA tónica**. Valor medio de los desempeños medios obtenidos durante la realización de 300 simulaciones de la tarea de selección forzada, cada una considerando aprendizaje durante 240 *trials* consecutivos.

Peso	$w^{Ctx_{cog}}_{Str_{cog}}[1]$	$w^{Ctx_{cog}}_{Str_{cog}}[2]$	$w^{Ctx_{cog}}_{Str_{cog}}[3]$	$w^{Ctx_{cog}}_{Str_{cog}}[4]$
Valor $[media\,(\pm)\,SD]$	0.6155 ± 0.0015	0.5326 ± 0.0014	0.4679 ± 0.0013	0.4362 ± 0.0063

Tabla 3.2: **Pesos cortico-estriados cognitivos post-aprendizaje**. Valores medios y desviación estándar estimados para cada figura k los respectivos pesos cortico-estriados del lazo cognitivo $w^{Ctx_{cog}}_{Str_{cog}}[k]$ luego de realizar 240 *trials* consecutivos, estimados utilizando un total de 300 realizaciones del experimento.

Se realiza un total de 600 *trials* por cada experimento, para cada nivel de DA, donde 100 *trials* presentan el mismo par de figuras de seis combinaciones posibles:

$$\{(0,1),\ (0,2),\ (0,3),\ (1,2),\ (1,3),\ (2,3)\}$$

Este experimento es repetido 100 veces, de las cuales se obtienen muestras del proceso de selección que permiten estimar valores medios de entropía.

Los valores de entropía sobre la selección de cada par g se calculan siguiento la ecuación (3.3.4), donde la probabilidad de seleccionar la figura cue_c dada la presentación del par g, $p(cue_c|g)$, se estima como el valor medio de la proporción de *trials* en los cuales se realiza esta selección sobre el total de repeticiones del experimento.

$$H(g) = -\sum_{c=1}^{2} p(cue_c|g)\ ln(p(cue_c|g)), \tag{3.3.4}$$

Dado que la presentación de los pares es un proceso determinístico, la única incerteza en este experimento proviene de qué figura es seleccionada en cada *trial*. En base a esto, en la ecuación (3.3.5) se define la entropía total $H(T)$ como la entropía media para cada par sobre todos los niveles de DA considerados. Se considera el cálculo de la media, en comparación a la suma, con el objetivo de tener una representación de la entropía total en una escala comparable al rango presente en las entropías de cada par $H(g)$.

$$H(T) = -\frac{1}{6}\sum_{g=1}^{6}\sum_{c=1}^{2} p(cue_c|g)\ ln(p(cue_c|g)) \tag{3.3.5}$$

Los resultados son presentados en la figura 3.9, donde se aprecian relaciones relaciones entre los niveles de DA tónica y las probabilidades estimadas de selección (figura 3.9 A), y en la respetiva representación en base a medidas de entropía (figura 3.9 b).

3.4. Discusión de los resultados

Los resultados obtenidos en *trials* individuales, presentados en la figura 3.5 demuestran un funcionamiento adecuado, en función de las condiciones descritas en la sección 3.3.2. El modelo propuesto de múltiples lazos cortico-ganglios basales con integración de DA tónica es capaz de realizar selecciones en todo el rango de DA utilizado. En estos experimentos se aprecia una relación entre el nivel de SNR y el nivel de DA tónica, donde a menor nivel de DA el ruido de entrada en el estriado tiene una mayor presencia en la actividad cortical.

Durante la fase de aprendizaje a través de *trials* consecutivos, independiente del nivel de DA el sistema aprende los mismos niveles de valores V_k (ver figura 3.7), los que son coherentes con la probabilidad de recompensa asociada a cada figura (ecuación (3.3.2)). Del mismo modo, los pesos cortico-estriados cognitivos resultantes (ver figura 3.8) presentan una evolución y valores finales similares (ver tabla 3.2). Por lo tanto, la integración de DA tónica en el modelo propuesto no altera los procesos internos de aprendizaje. Sin embargo, el desempeño sí se ve modulado por el nivel de DA (ver figura 3.6), por lo que si bien el sistema aprende, cuánto se considera este aprendizaje sobre los procesos de selección se ve regulado por el nivel de DA tónica.

Finalmente, los datos de *trials* individuales post-aprendizaje presentados en la figura 3.9 comprueban que, aunque los pesos cortico-estriados son equivalentes entre distintos niveles de DA, la probabilidad de seleccionar las figuras correctas aumenta si el nivel de DA es mayor, presentando un comportamiento con menor razón de exploración-explotación: aumentos en los niveles de DA tónica disminuyen la entropía asociada a los procesos de selección, presentando tendencias más explotativas sobre el conocimiento adquirido.

El único par para el cuál no se aprecian cambios en las probabilidades de selección (figura 3.9) es el par $g = [2, 3]$, correspondiente a las figuras con menor probabilidad de recompensa. Este efecto se encuentra directamente relacionado al bajo nivel en las respectivas conexiones cortico-estriadas post-aprendizaje (ver tabla 3.2), cuya diferencia no es suficiente para ser explotada por los niveles de DA tónica.

En síntesis, en base a los resultados expuestos, niveles menores de DA tónica disminuyen el SNR en las entradas sinápticas del estriado, aumentando las probabilidades de que el ruido guíe el proceso de selección, aumentando la exploración de alternativas consideradas como menos beneficiosas. Al contrario, niveles mayores de DA tónica aumentan el SNR, derivando a una mayor probabilidad de que la selección se lleve a cabo en función de los niveles de los pesos cortico-estriados.

Por lo tanto, en el modelo propuesto los niveles de DA tónica modulan el comportamiento de selección en términos de presentar mayores niveles de exploración de las alternativas disponibles, o bien, de presentar un comportamiento con mayores tendencias de explotar conocimiento previamente adquirido.

3.5. Conclusiones

Se presenta un modelo de múltiples lazos cortico-ganlgios basales con integración de efectos de los niveles de DA tónica en el estriado. Esta implementación considera efectos en las conexiones cortico-estriadas y en la función de transferencia del estriado en ambos lazos, cognitivo y motor. Además, los lazos del modelo mantienen una simetría en términos de estructura de la red, característica asociada a los modelos de múltiples lazos cortico-ganglios basales, donde cada lazo comprime un módulo idéntico de selección de acciones.

Resultados obtenidos en tareas de selección individual, en tareas de selección consecutivas considerando aprendizaje, y en tareas de evaluación del modelo en un estado post-aprendizaje, demuestan que la implementación propuesta mantiene un correcto funcionamiento del modelo con respecto a procesos de selección y aprendizaje, además de integrar una modulación de la razón entre exploración y explotación basada en niveles de DA tónica. La modulación obtenida es coherente con resultados previos, donde el nivel de DA D1 tónica escala las diferencias entre los niveles de actividad cortical de entrada al estriado (directamente proporcionales a los pesos cortico-estriados) [13]. Por ello, a través de la realización de una tarea de aprendizaje, una vez separados los niveles de los pesos cortico-estriados cognitivos, mayores niveles de DA tónica promueven la explotación de estas diferencias, reduciendo a su vez selecciones exploratorias.

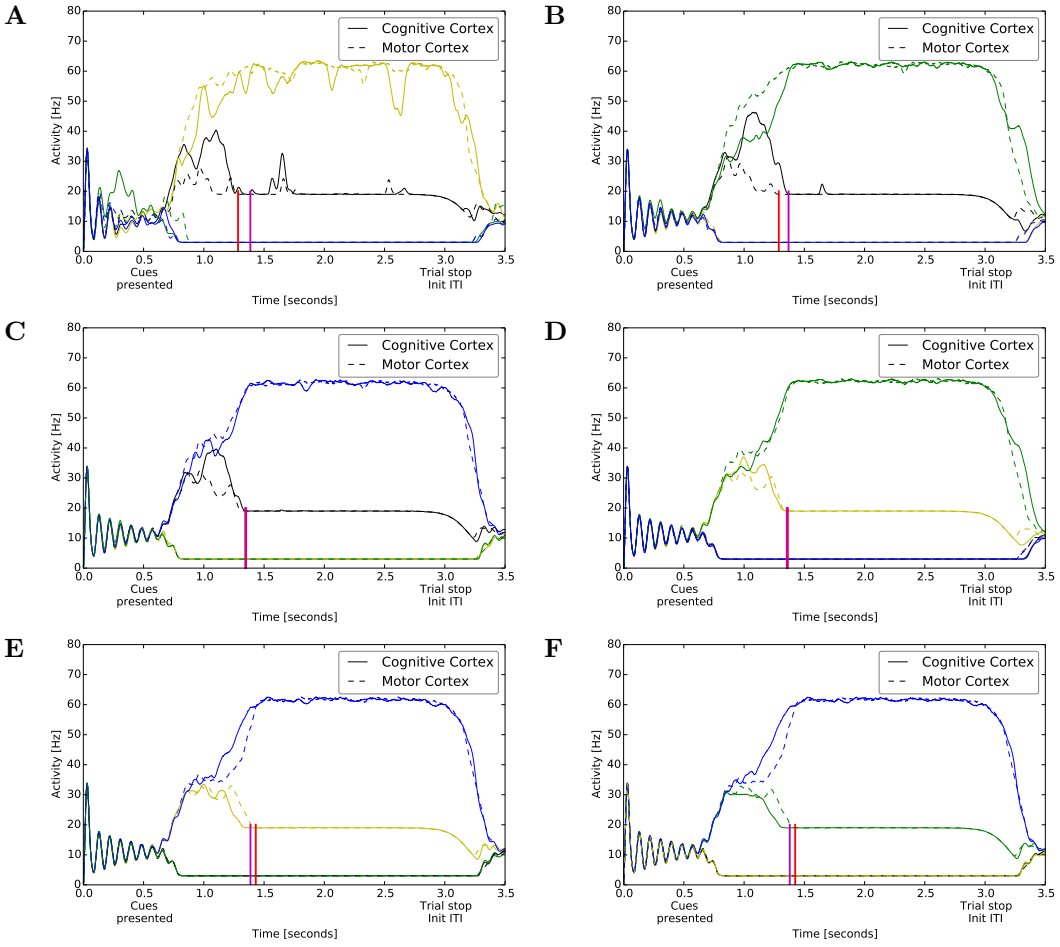

Figura 3.5: **Efectos del nivel de DA tónica en la realización de *trials* individuales en una tarea de selección forzada de dos opciones** Actividad cortica cognitiva y motora para cada nivel de DA tónica evaluado, con $DA \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5\}$. Las líneas contínuas corresponden a actividad cortical del lazo cognitivo, mientras que las líneas intermitentes presentan la actividad cortical del lazo motor. Se presentan 4 colores (amarillo, verde, azul y negro), cada uno representando una *alternativa* distinta. Las señales del lazo motor, las *acciones*, utilizan el color respectivo de la *alternativa* asociada. El modelo es capaz de llevar a cabo una selección entre las dos figuras presentadas, para todos los niveles de DA. Después de la presentación ($t > 2500$ $[ms]$), las actividades sinápticas regresan al nivel de estabilización alcanzado previa a la presentación de las figuras ($t < 500$ $[ms]$). Las líneas verticales moradas y rojas corresponden a los tiempos en los que se marca una selección a nivel cognitivo y motor, respectivamente. **A** Nivel de DA igual a 0.0, con umbral en el estriado cognitivo y motor igual a 18.5 $[sp/s]$. **B** Nivel de DA igual a 0.5 con un umbral en el estriado $\{cog, mot\}$ de 29.0 $[sp/s]$. **C** Nivel de DA igual a 1.0 con un umbral en el estriado $\{cog, mot\}$ de 39.5 $[sp/s]$. **D** Nivel de DA igual a 1.5 con un umbral en el estriado $\{cog, mot\}$ de 50.0 $[sp/s]$. **E** Nivel de DA igual a 2.0 con un umbral en el estriado $\{cog, mot\}$ de 60.5 $[sp/s]$. **F** Nivel de DA igual a 2.5 con un umbral en el estriado $\{cog, mot\}$ de 71.0 $[sp/s]$.

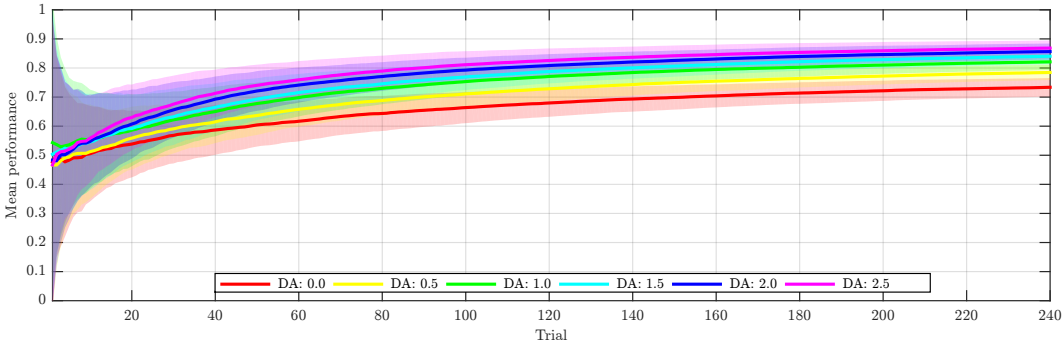

Figura 3.6: **Effectos de DA tónica en el desempeño de una tarea de selección forzada de dos opciones.** Para cada nivel de DA, se prueba el modelo en una tarea de selección forzada con dos opciones considerando aprendizaje entre *trials* (ver sección 3.3.1). Se presenta el desempeño medio, definido en la ecuación (3.3.3). La tarea es reproducida 300 veces para cada nivel de DA aplicado, donde cada tarea considera 240 *trials* consecutivos en los que se realiza una selección. Se descartan *trials* intermedios sin selección, para los que el sistema no cumple con la diferencia mínima en los niveles de actividad cortical. Se presenta el valor medio (líneas sólidas) con la correspondiente desviación estándar (sombras), calculados sobre el total de reproducciones. Se aprecia que cambios en el nivel de DA tónica tiene efectos coherentes en el desempeño medio alcanzado al finalizar la tarea: para un mayor nivel de DA se alcanza un mayor nivel de desempeño medio, para la tarea evaluada.

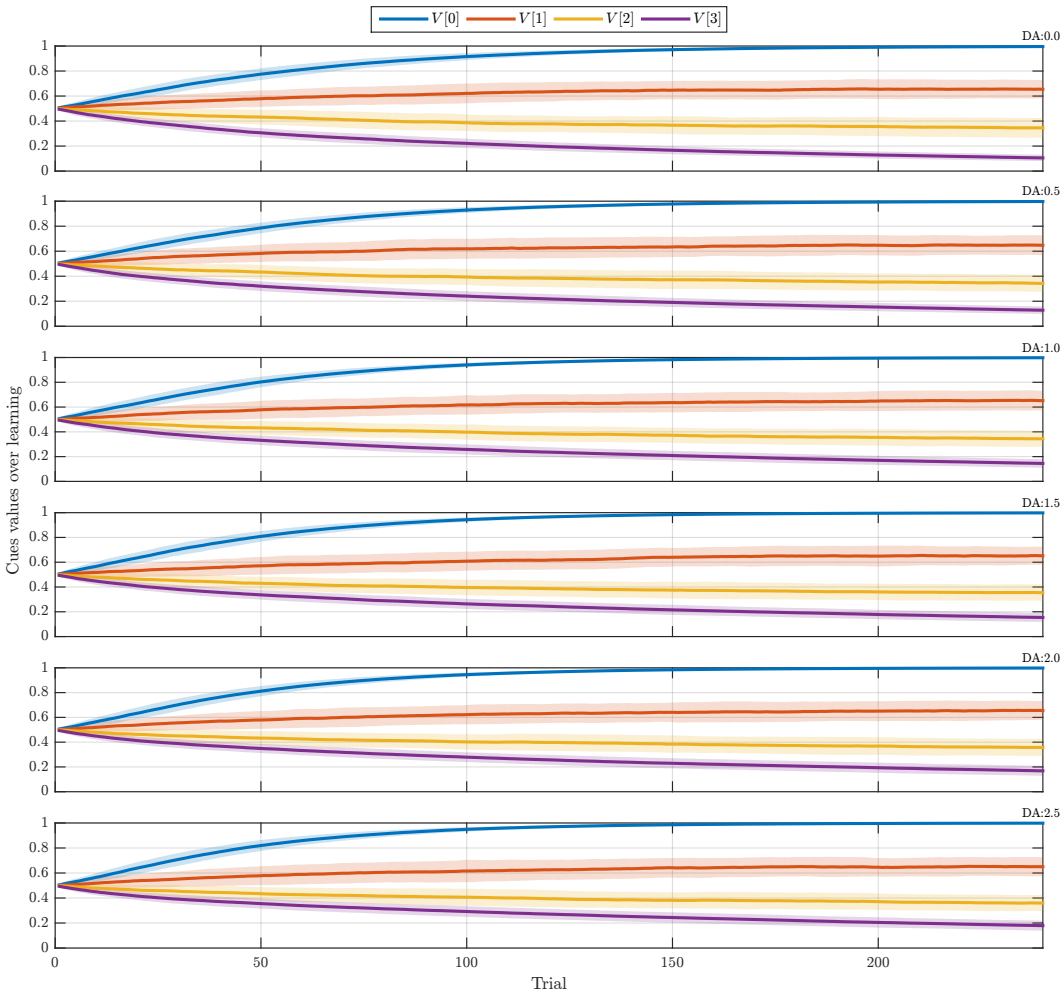

Figura 3.7: **Valor interno de las figuras a través del aprendizaje.** Evolución de V_k, valores internos asociados a cada *alternativa* k, los cuales simulan memoria de largo plazo los pesos sinápticos de las conexiones cortico-estriadas cognitivas. Para cada nivel de DA evaluado (de *arriba* hacia *abajo*: $DA = 0.0$, $DA = 0.5$, $DA = 1.0$, $DA = 1.5$, $DA = 2.0$ y $DA = 2.5$), se presenta el valor medio (líneas sólidas) con cada desviación estándar correspondiente (sombras). Tanto el comportamiento de la evolución de los valores como los niveles alcanzados una vez completado el aprendizaje son similares, lo que indica que no existe una relación entre el nivel de dopamina y el aprendizaje de los valores V_k.

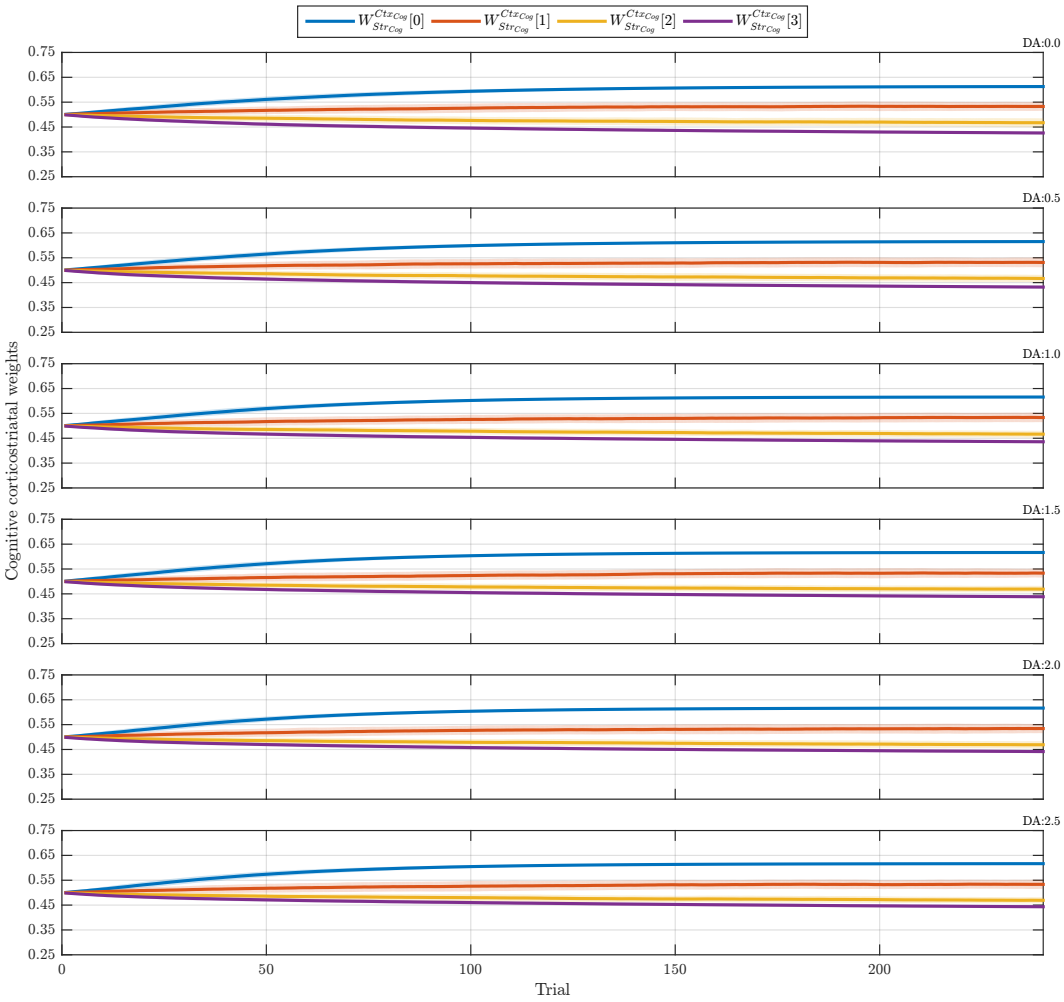

Figura 3.8: **Pesos cortico-estriados cognitivos a través del aprendizaje.** Evolución de los pesos sinápticos de las conexiones cortico-estriadas cognitivas. Se presentan los pesos para cada nivel de DA evaluado, presentando desde *arriba* hacia *abajo*: $DA = 0.0$, $DA = 0.5$, $DA = 1.0$, $DA = 1.5$, $DA = 2.0$ y $DA = 2.5$. Se presenta el valor medio (líneas sólidas) con cada desviación estándar (SD) correspondiente (sombras). Es apreciable la similitud en la evolución de los pesos, cuyos valores finales son aproximablemente iguales (ver tabla 3.2). Esto indica que la DA tónica no afecta el aprendizaje en los pesos cortico-estriados.

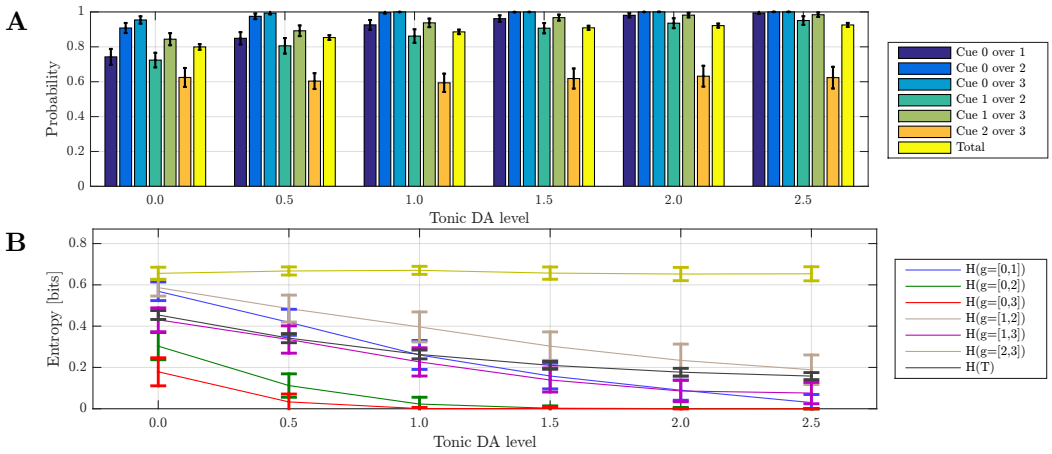

Figura 3.9: **Efectos del nivel de DA tónica en selecciones post-aprendizaje.** Evaluación de la probabilidad de selección de cada figura, para cada par de figuras posibles. Se presentan valores medios computados sobre 100 realizaciones del experimento descrito en la sección 3.3.4, cada uno consistente en 100 presentaciones para cada par de figuras (600 en total). Las barras de error presentadas corresponden a la respectiva desviación estándar. **A** Estimación de la probabilidad de seleccionar, para cada par posible, la figura asociada a una mayor probabilidad de obtener recompensa. La estimación de la probabilidad total es calculada considerando selecciones para sobre todos los pares. Barras presentando un valor mayor a un 50 % demuestran un proceso de aprendizaje exitoso. Incrementos en los niveles de las barras, en función de los niveles de DA tónica, muestran cómo el sistema presenta una mayor tendencia a seleccionar las figuras con mayor probabilidad de recompensa. **B** Entropía asociada a las estimaciones de la probabilidad de selección, calculadas en función de las ecuaciones (3.3.4) y (3.3.5). Esta representación, utilizando medidas de entropía, muestra en forma clara que para un mayor nivel de DA tónica el sistema es menos variable (menor exploración, mayor tendencia a explotar el conocimiento adquirido) con respecto al proceso de selección, con menores niveles de entropía.

Capítulo 4

CONTROLADOR BIO-INSPIRADO: IMPLEMENTACIÓN DEL MODELO PROPUESTO COMO MECANISMO DE TOMA DE DECISIONES

4.1. Introducción

Para un agente autónomo la exploración es esencial, ya que permite aprender y mejorar el comportamiento en relación a resultados esperados. Sin embargo, también expone al agente a situaciones desconocidas, potencialmente dañinas. Por ello, debe existir una razón adecuada entre exploración-explotación [4].

Si bien desde la teoría del Aprendizaje Reforzado (RL) clásico se define una aproximación al control de la razón de exploración-explotación a través del parámetro conocido como la temperatura inversa β [53], el control de este parámetro es uno de los problemas teóricos más importantes del área [23]. Una alternativa a esta implementación clásica es el definir un controlador que utilice el modelo de lazos CBG como mecanismo de selección de acciones, permitiendo controlar las razones de exploración-explotación. Este controlador puede ser extendido posteriormente para incluir lazos de control sobre los niveles de DA tónica basado en mecanismos biológicos.

La integración de los lazos CBG en un controlador de una plataforma robótica permite:

- Evaluar el modelo de lazos CBG como mecanismo de selección de acciones. Para ello, se simula una tarea simple de supervivencia.
- Evaluar los efectos de cambios en los niveles de DA en términos de desempeño, y corroborar sus efectos en términos de la tasa de exploración-explotación en una tarea más compleja a la implementada en el capítulo 3.

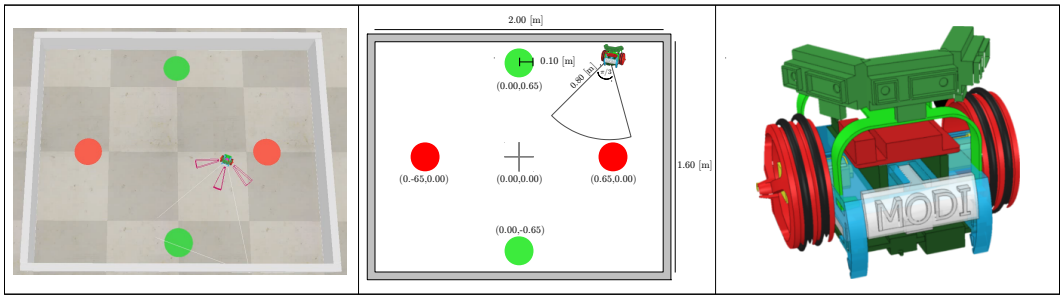

Figura 4.1: **Escenario virtual de la tarea de supervivencia.** El escenario simulado es un ambiente cerrado de forma rectangular, delimitado por paredes de 20 [cm] de altura. Los cuatro círculos pintados a nivel de piso, dos rojos y dos verdes, corresponden a fuentes de energía vital y potencial, respectivamente. ***Izquierda*** Escenario virtual, en el simulador V-REP. Las figuras moradas que rodean al MODI corresponden a las zonas de detección de los sensores de proximidad añadidos al modelo virtual. ***Centro*** Medidas del escenario, incluyendo en el diagrama los límites del rango de percepción virtual implementado en el robot MODI. ***Derecha*** Robot MODI simulado.

A continuación, en la sección 4.2, se detalla la tarea de supervivencia implementada, presentando las especificaciones del ambiente y las características del agente autónomo consideradas para enfrentar la tarea. En la sección 4.3 se presenta el controlador implementado, integrando el modelo de los lazos cortico-ganglios basales (CBG) como mecanismo de toma de decisiones. Posteriormente, en la sección 4.4 se describen y evalúan los resultados obtenidos sobre los efectos del nivel de dopamina en el comportamiento del agente. El capítulo finaliza con breves conclusiones descritas en la sección 4.5.

4.2. Tarea de supervivencia de dos recursos

Con el objetivo de poner a prueba el modelo descrito en el capítulo 3 como un mecanismo de selección de acciones en un robot autónomo, se implementa una tarea simple de supervivencia de dos recursos. Este tipo de tareas corresponde al mínimo escenario para probar mecanismos de selección de acciones [50], y permite caracterizar el comportamiento del sistema enfrentando una tarea que requiere constante aprendizaje, ya que las necesidades del agente artificial varían en el tiempo, cambiando las reglas de recompensa.

4.2.1. Descripción general

La tarea considera la existencia de dos tipos de energía, recursos de supervivencia, en un agente simulado. El primer tipo de energía corresponde a energía vital VE, que es permanentemente consumida a través del tiempo y que debe mantenerse siempre con un valor positivo: nivel de $VE = 0$ representa la muerte del agente. Para incrementar el nivel de energía vital, el agente debe transformar energía potencial PE, correspondiente al segundo tipo de energía, de forma similar al consumo de alimento.

El espacio donde se desenvuelve el agente es un ambiente cerrado, delimitado por paredes, que contiene dos fuentes de energía para cada tipo. El agente debe posicionarse sobre una de estas fuentes de energía para poder incrementar el nivel de energía respectivo. Entonces, para

mantenerse con vida, el agente debe adquirir PE desde alguna fuente de energía potencial PE_s, para luego transformar esta energía en VE, estando posicionado sobre una fuente de energía vital VE_s.

El escenario simulado es presentado en la figura 4.1, donde las fuentes de energía se simulan mediante círculos de color ubicados en el suelo (rojo para VE_s y verde para PE_s). La posición de las fuentes es constante, y simétrica con respecto al centro del escenario.

Como software de simulación gráfica y de computación de interacciones físicas, se utiliza el simulador V-REP (*Virtual Robot Experimentation Platform*) [10], configurado para utilizar la biblioteca *Bullet Physics Library*[1], ya incorporada en el simulador. Se utiliza V-REP debido a que presenta una serie de bibliotecas tipo API (Interfaz de Programación de Aplicaciones) que permiten controlar la funcionalidad del simulador desde un cliente remoto, disponibles para varios lenguajes de programación diferentes, entre ellos *C++*, *MATLAB* y *Python*. Esto entrega libertad con respecto al lenguaje y paradigma de programación, permitiendo utilizar una extensión de la implementación computacional del modelo de lazos cortico-ganglios basales previamente desarrollada en *Python*.

4.2.2. Capacidades del agente simulado

Dada las características de la tarea de supervivencia, y la naturaleza cognitiva y motora de los lazos cortico-ganglios basales utilizados como mecanismo de selección de acciones, se definen las capacidades del agente simulado como un conjunto de *acciones* motoras y un conjunto de *alternativas* asociadas a una decisión cognitiva.

En términos de percepción sensorial, el agente es dotado con las capacidades de detectar paredes, y de detectar fuentes de energía dentro de un rango de visión virtual, capacidades descritas en detalle en la sección 4.2.5.

Con respecto al conjunto de *acciones* considerado, el agente es capaz de ejecutar cinco movimientos motores:

m_i $forward$: movimiento en dirección frontal.

m_{ii} $turn\_left$: giro hacia la izquierda sobre su propio eje.

m_{iii} $turn\_right$: giro hacia la derecha sobre su propio eje.

m_{iv} *stop*: el agente se detiene para reducir la tasa de consumo de energía vital a la mitad.

m_v *reload*: recarga de algún tipo de energía, estando posicionado sobre una fuente de energía respectiva.

De forma análoga, se define un conjunto de cinco *alternativas* cognitivas para el agente:

c_i $Wander$: con el objetivo de explorar el ambiente, el agente ejecuta una *acción* seleccionada de forma aleatoria entre $forward$, $turn\_left$ y $turn\_right$.

c_{ii} *Rest*: simula la selección de descansar, ejecutando el movimiento *stop*.

[1]Biblioteca de Detección de Colisiones y Dinámica de cuerpos rígidos, disponible como Software libre en http://www.bulletphysics.org.

c_{iii} $Wall_{av}$: evasión de colisiones con paredes. Dependiendo de las posición donde se dectecta la pared, el agente ejecuta un movimiento un movimiento de giro hacia la dirección contraria.

c_{iv} $Reload_{VE}$: el agente actúa con el objetivo de incrementar su nivel de VE. Para ello, dependiendo de la posición del agente con respecto a una fuente VE_s, ejecuta movimientos para reducir la distancia con la fuente ó, una vez estando suficientemente cerca, recargar transformando PE en VE.

c_v $Reload_{PE}$: de forma equivalente a $Reload_{VE}$, el agente actúa con el objetivo de incrementar su nivel de PE, ya sea acercándose a una fuente PE_s ó, estando posicionado sobre la fuente, recargando.

4.2.3. Tasas de consumo de los niveles de energía vital y potencial

Ambos niveles de energía, VE y PE, se definen dentro del rango $[0, 1]$. La energía vital del agente es consumida de forma inherente a una tasa constante, excepto cuando el agente se encuentra ejecutando *Rest*, caso en el que la tasa de consumo se ve reducida. Siempre que el agente se encuentre posicionado sobre una fuente VE_s, es capaz de incrementar su nivel de VE a costa de consumir PE. A su vez, estando posicionado sobre una fuente PE_s, el agente es capaz de incrementar su nivel de PE sin costo.

Por lo tanto, dependiendo del par $\{alternativa, acción\}$ seleccionado, las tasas de consumo de energía varían según:

$$\Delta VE = \begin{cases} \alpha_{PE} & \text{si } \{Reload_{VE}, reload\} \\ 0.5 \cdot \alpha_{VE} & \text{si } \{Rest, stop\} \\ \alpha_{VE} & \text{Cualquier otro caso} \end{cases}$$

$$\Delta PE = \begin{cases} -\alpha_{PE} & \text{si } \{Reload_{VE}, reload\} \\ \alpha_{PE} & \text{si } \{Reload_{PE}, reload\} \\ 0 & \text{Cualquier otro caso} \end{cases}$$

Los parámetros α_{VE} y α_{PE} se definen de forma tal que el mínimo tiempo de vida del agente sea de 80 $[s]$, con un tiempo de carga máximo de 5 $[s]$, considerando niveles iniciales $VE_0 = 0.8$ y $PE_0 = 0.2$. Todas las constantes aplicadas en el modelo se detallan en la sección A.1.

4.2.4. Definición de condiciones de recompensa

Para definir cuándo las *alternativas* serán recompensadas, se establecen tres tipos de comportamiento posibles para el agente:

- *PE seeker*: corresponde al comportamiento de mayor prioridad, activado en caso de presentarse niveles bajos de PE. Mientras el agente se encuentre en este comportamiento, las recompensas son determinadas en función si es que la *acción* ayudó de alguna forma a aumentar el nivel de PE.

- *VE seeker*: activado en caso de presentarse niveles de VE bajo el nivel medio, siempre y cuando no se cumpla la condición de activación del comportamiento anterior. Mientras el agente se encuentre en un comportamiento tipo *VE seeker*, las recompensas son determinadas en función de si es que la *acción* ayudó de alguna forma a aumentar el nivel de PE.
- *Both*: activado en caso de no cumplirse las condiciones de activación del resto de los comportamientos. Mientras el agente se encuentre en un comportamiento tipo *both*, las recompensas son determinadas en función si es que la *acción* ayudó de alguna forma a aumentar cualquier nivel de energía.

Las condiciones se describen en detalle en la tabla 4.1. Estas condiciones fueron determinadas de forma arbitraria, estableciendo una guía para el agente para enfrentar la tarea de supervivencia.

4.2.5. Plataforma robótica MODI

La plataforma robótica utilizada como agente es el robot MODI (MODular Intelligent) [44], plataforma móvil compacta de dos ruedas, de tipo diferencial y de hardware abierto[2], con capacidades de comunicación inalámbrica diseñada para aplicaciones de enjambes de robots. El diseño de esta plataforma no considera sensores. La versión virtual utilizada es presentada en la figura 4.1 C.

Para añadirle las capacidades de detección de paredes y de detección de fuentes de energía consideradas para el agente (ver sección 4.2.2), se integran al MODI sensores de proximidad y un rango de visión virtual para detección de fuentes. Se instalan 3 sensores de proximidad con un rango de detección entre 4 y 16 $[cm]$, posicionados a 3 $[cm]$ del centro del robot, y a 6 $[cm]$ del suelo, separados por $\pi/6$ $[rad]$ con respecto del sensor central (orientado hacia el frente).

[2]Diseño disponible en https://github.com/mjescobar/MODI.

Tabla 4.1: **Condiciones de recompensa presentadas en orden de prioridad decendiente.**

Comportamiento	Rango de activación	Condición de recompensa
PE seeker	Nivel de PE igual o menor a 0.2	¿Se encuentra el robot más cerca de PE_s (en términos de posición u orientación)?, o, ¿el nivel de PE aumentó?
VE seeker	Nivel de VE igual o menor a 0.5	¿Se encuentra el robot más cerca de VE_s (en términos de posición u orientación)?, o, ¿el nivel de VE aumentó?
Both	Otros casos	¿Se encuentra el robot más cerca de alguna fuente de energía?, o, ¿el nivel de VE o PE aumentó?

La percepción considerada para detectar las fuentes de energía consiste en que el agente conoce de forma instantánea la posición de una fuente de energía cuando ésta se encuentra posicionada dentro de un rango de visión determinado por un arco (ver figura 4.1 B). Esta implementación puede ser reemplazada, por ejemplo, por la utilización de una cámara frontal y la aplicación de algoritmos de segmentación como el presentado en [7].

4.3. Controlador

El controlador de la paltaforma robótica es implementado en forma de una máquina de estados finita de tipo determinista compuesta por cuatro estados, definidos como etapas: *Percibir*, *Decidir*, *Ejecutar* y *Evaluar*.

En cada instante de tiempo discreto, se lleva a cabo una iteración del lazo del controlador: se inicia el lazo en el estado *Percibir*, continuando con las respectivas reglas de transición, descritas a continuación. Al final de cada iteración, se envía una señal de control al simulador V-REP que dispara el cómputo de un paso de tiempo discreto de una duración igual a $dt_{VREP} = 50$ milisegundos. La figura 4.2 presenta un diagrama del controlador implementado.

4.3.1. *Percibir*

Durante esta etapa el robot sensa el ambiente y percibe su propio estado (posición, orientación y niveles de energía), define las *alternativas* y *acciones* disponibles, y configura los lazos CBG en caso de ser necesario.

El sensado del ambiente considera la detección de paredes basado en mediciones de los sensores de proximidad, la detección de fuentes de energía y, con esta información, los pares {*alternativa*,*acción*} presentes.

En todo momento, solo dos pares {*alternativa*,*acción*} se encuentran disponibles de forma simultánea para el robot, en función de las características del ambiente sensadas. Esta restricción se aplica para mantener los parámetros del modelo de lazos CBG descrito en el capítulo 3, ya que aumentar el número de pares disponibles modifica el punto de operación del modelo, inhabilitando el proceso de selección (datos no incluídos). Las relaciones entre características sensadas y pares disponibles se persentan en la tabla 4.2.

Al definir cada *acción* disponible, se definen cotas para su actuación, asociando un límite de distancia o grados de rotación cuando corresponda, y un tiempo máximo de ejecución definido de forma pseudo-aleatoria como $t_{max} = 1.5 + u \sim \mathcal{U}(-0.4, 0.4)$ $[s]$.

Se considera que durante un tiempo mínimo, equivalente a tres instantes discretos, se deben detectar las mismas *alternativas* para que el robot las procese como disponibles. Esto evita indecisiones en cuanto a las *alternativas* disponibles por detecciones instantáneas, otorgando un nivel de histéresis al sistema asociable a periodos requeridos por procesos de percepción para convertir esta situación sensada a un nivel consciente.

Si es que se ha detectado un cambio en las *alternativas* disponibles, o se ha finalizado una *acción* en ejecución, el lazo CBG se lleva a un estado inicial. Posteriormente, se configuran las entradas corticales del lazo CBG en función de los pares {*alternativa*,*acción*} disponibles. Una vez hecho esto, se establece que el robot se encuentra decidiendo. En el caso específico

Figura 4.2: **Diagrama del controlador de la plataforma, integrando el modelo de lazos CBG como mecanismo de selección.** El controlador es implementado como un proceso iretarivo compuesto por cuatro etapas: *Percibir*, *Decidir*, *Ejecutar* y *Evaluar*. Cada etapa es descrita en detalle en la sección sección 4.3. Las flechas naranjas muestran todas las transiciones posibles durante una iteración, cada una asociada a un motivo del por qué esa transición es tomada (cuadrados verdes). Cada transición comienza en un cuadrado naranja, terminando en una flecha cóncava. Una iteración comienza en la etapa *Percibir*, la cual puede conducir a la etapa *Decidir* si es que se detectan nuevas *alternativas* disponibles o si el agente se encuentra decidiendo, o a la etapa *Evaluar* (ruta naranja intermitente entre dos círculos) para, una vez terminado su proceso, continúar desde la etapa *Percibir* (ruta naranja intermintente entre un círculo y un cuadrado). Desde la etapa *Decidir* es posible continuar comenzando la ejecución de una *acción* si es que se ha realizado una selección, o de mantenerse decidiendo, finalizando la iteración. De forma similar, desde la etapa *Ejecutar* el robot puede continuar con la evaluar de una *acción* finalizada, o mantenerse actuando, finalizando la iteración. Después de haber realizado una evaluación, durante la etapa *Evaluar* (desde la etapa *Ejecutar*), se termina la iteración actual. Al finalizar cada iteración se dispara el proceso de simulación para un paso de tiempo.

de haberse detectado un cambio de *alternativas*, se fuerza una *evaluación* intermedia de forma transitoria para determinar si la *alternativa* asociada a la *acción* en ejecución fue beneficiosa o nó (proceso descrito en detalle en la sección 4.3.4).

Si es que se encuentra una *acción* en ejecución, se transita a la etapa *Ejecutar*. Del mismo modo, si el robot se encuentra decidiendo a través de los lazos CBG, se transita a la etapa *Decidir*.

4.3.2. *Decidir*

Cada vez que se ingresa a esta etapa se simula un paso de tiempo en los lazos CBG. Si es que se realiza una selección, se determina que el robot se encuentra en ejecución de una *acción*, guardando datos el estado presente del robot, y una marca de tiempo. La etapa a continuación. en caso de encontrarse en ejecución, es *Ejecutar*. En caso contrario, se termina el lazo del controlado volviendo al estado inicial, *Percibir*.

4.3.3. *Ejecutar*

Durante esta etapa, el robot lleva a cabo, en el tiempo, una *acción* previamente seleccionada. Si es que alguno de los límites asociados a la *acción*, determinados al haberse establecido la *acción* como disponible (ver sección 4.3.1), se determina que la ejecución de la *acción* ha finalizado, en cuyo caso se transita a la etapa *Evaluar*. En caso contrario, se termina el lazo del controlado volviendo al estado inicial.

4.3.4. *Evaluar*

Esta etapa es alcanzable sólo en caso de que se realize una transición de forma transitoria desde la etapa *Percibir*, o una vez que se ha completado la ejecución de una acción en la etapa *Ejecutar*. El robot lleva a cabo el análisis de si la *acción* ejecutada fue realmente beneficiosa, computando una recompensa en función de los objetivos actuales del comportamiento del robot siguiendo las condiciones descritas en la sección 4.2.4. La información sobre si la *acción* derivó o nó a una recompensa dispara el proceso de aprendizaje sobre el lazo CBG cognitivo.

En caso de haber alcanzado este estado desde la etapa *Percibir*, se continúa considerando las reglas de transición de esa etapa. Al contrario, si se ha accedido desde la etapa *Ejecutar*, se cierra el lazo del controlador volviendo al estado inicial.

4.4. Resultados

Para cada nivel de DA, se realizan 300 simulaciones de la tarea de supervivencia implementada. Para cada simulación, se determina de forma pseudo-aleatoria la posición y orientación inicial del robot MODI. Cada simulación finaliza si es que el nivel de VE llega a cero o si el tiempo de simulación alcanza los 900 $[s]$.

En la figura 4.3 se presentan los pares $\{alternativa, acción\}$ disponibles para el robot en el tiempo, presentando instantes en los que el robot se encuentra decidiendo, y los respectivos instantes en los que se encuentra en ejecución una *acción* seleccionada.

Dada la tarea, como se implementa un controlador con el objetivo de mantenerse con vida, se realiza un análisis de las simulaciones en términos de la duración de éstas, equi-

Tabla 4.2: ***Alternativas* disponibles para el robot, en base a características del entorno**

Detección			***Alternativas* disponibles**	
Pared	VE_s	PE_s	Alternativa 1	Alternativa 2
0	0	0	$Wander$	$Rest$
0	0	1	$Wander$	$Reload_p$
0	1	0	$Wander$	$Reload_v$
0	1	1	$Wander$	$Reload_p$ o $Reload_v$ (aleatorio)
1	0	0	$Wander$	$Wall_{av}$
1	0	1	$Wall_{av}$	$Reload_P$
1	1	0	$Wall_{av}$	$Reload_V$
1	1	1	$Wall_{av}$	$Reload_p$ o $Reload_v$ (aleatorio)

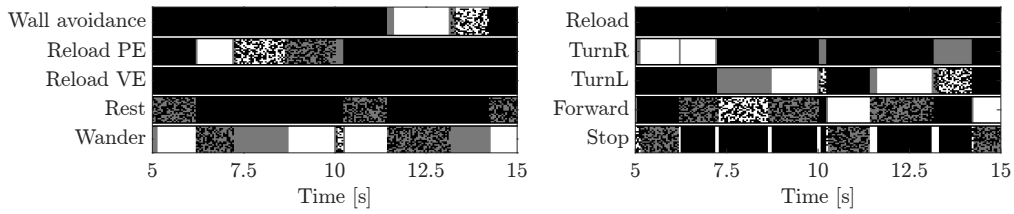

Figura 4.3: **Ejemplo de disponibilidad y selección de *alternativas* y *acciones*.** Considerando un nivel de $DA = 0.0$, se presenta una porción de un *trial* ejecutando la tarea de supervivencia descrita en la sección 4.2. Para diferenciar cada par {*alternativa,acción*}, se presenta un par en colores planos y otro par en colores granulados. Las opciones disponibles se muestran en gris, mientras que las mostradas en blanco corresponden a opciones en ejecución.

valentes al tiempo de vida alcanzado por el robot. Los resultados se presentan en la figura 4.4.

La figura 4.4 A muestra histogramas del tiempo de vida alcanzado para cada nivel de DA. Mientras que el nivel de DA tónica aumenta, disminuye la cantidad de simulaciones de corta duración (menor a [200 [s]), y aumenta la cantidad de simulaciones con una duración igual al máximo posible (900 [s]). Únicamente comparando los dos niveles de DA más altos $\{1.5, 2.0\}$, se invierte esta tendencia. Al calcular la media en los histogramas se obtiene una estimación de la esperanza de vida respectiva para cada nivel de DA, donde se presenta un aumento coherente entre DA y el valor estimado de esperanza de vida, para $DA \in [0.0, 1.5]$, y una disminución al comparar entre $DA = 1.5$ y $DA = 2.0$. Estos resultados sugieren la existencia de un nivel específico de DA tónica tal que se maximiza la esperanza de vida, para la tarea y controlador implementados.

La figura 4.4 B presenta efectos del nivel de DA en valores medios de los porcentages de tiempo utilizados ejecutando cada *alternativa*. Para niveles de DA tónica dentro del rango $[0.0, 1.5]$, se presenta una reducción del procentage de tiempo utilizado en *acciones* asociadas a la *alternativa Wander*, mientras que el porcentage de tiempo utilizado en *acciones* asociadas a recargas de los niveles de VE y PE aumentan ($Reload_{VE}$ y $Reload_{PE}$, respectivamente). Entonces, para niveles de DA dentro del rango comentado, a medida que incrementa la DA el robot realiza menos acciones de exploración, presentando una mayor explotación ya que el robot presenta un mayor interés en realizar *acciones* asociadas a *alternativas* que incrementan los niveles energía.

La figura 4.4 C muestra trayectorias para 50 simulaciones para distintos niveles de DA tónica. Para mayores niveles de DA, el robot realiza un mayor número de trayectorias entre fuentes de energía. Esto demuestra una selección de acciones con mayor coherencia para niveles de DA mayores, i.e., menor entropía en los procesos de selección.

4.5. Conclusiones

El lazo de control propuesto aplicado en la tarea de supervivencia de dos recursos, mínimo escenario para evaluar mecanismos de toma de decisiones, es capaz de encontrar y seguir estrategias que resuelven el desafío. Por lo tanto, se determina que la integración del lazo CBG afectado por dopamina tónica, definido en el capítulo 3, como mecanismo de toma de

decisiones es correcta. Durante la resolución de la tarea, el robot aprende en tiempo real, en base a recompensas, las acciones que aumentan los beneficios esperados para el estado presente del robot (con respecto a sus niveles actuales de energía).

Los efectos de la dopamina tónica a nivel subcortical permiten regular el comportamiento del robot en términos de la razon entre exploración y explotación. Como consecuencia, la esperanza de vida del robot se ve directamente afectada por el nivel de dopamina. Incluso, los datos sugieren la existencia de un nivel constante de dopamina tónica tal que la esperanza de vida es máxima sobre todos los niveles de dopamina.

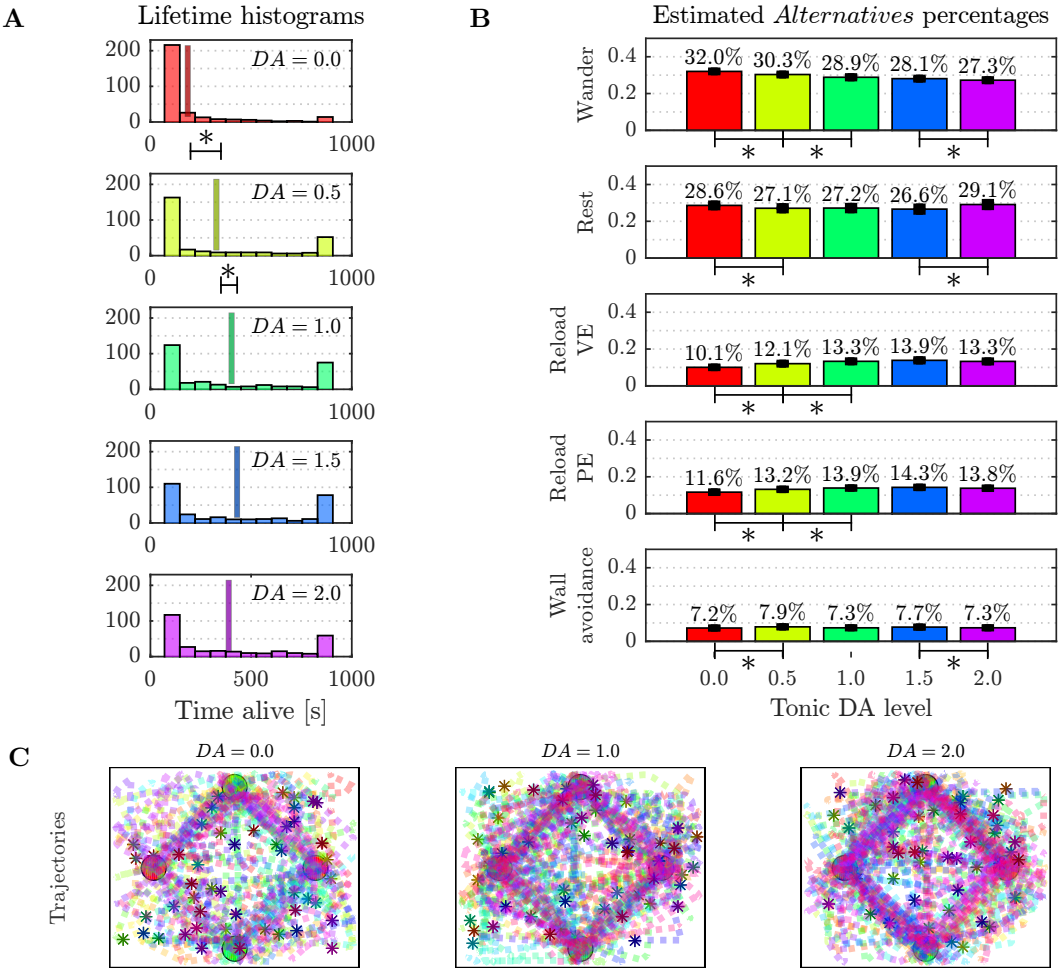

Figura 4.4: **Características del comportamiento del robot en la tarea de supervivencia de dos recursos.** Para cada nivel de DA tónica, se simulan 300 realizaciones de la tarea de supervivencia. Cada simulación finaliza si es que la energía del robot MODI decae a cero, siguiendo las dinámicas descritas en la sección 4.2.3 – con una configuración tal que el tiempo de vida mínimo es igual a 80 $[s]$–, o si se alcanza una duración de 900 segundos. La posición inicial del robot es determinada de forma aleatoria. **A** Histogramas del tiempo de vida alcanzado por el robot, presentados con intervalos de 100 $[s]$. Las líneas delgadas verticales indican la esperanza de vida media, calculada a partir de los datos del histograma respectivo. Los $*$ presentados entre gráficos indican que las distribuciones de los datos comparados presentan medias corridas a la derecha (indicado por la flecha bajo el símbolo) estadísticamente significativas ($p < 0.05$), siguiendo la prueba U de Mann-Whitney. **B** Valores medios de los porcentajes de tiempo utilizados en ejecutar cada *alternativa* para cada nivel de DA, donde los rectángulos negros sobre las barras presentan una altura igual al doble de la varianza. Los porcentages faltantes corresponden a periodos dedicados a decidir entre *alternativas* disponibles. Los $*$ indican movimientos estadísticamente significativos ($p < 0.05$), siguiendo la prueba U de Mann-Whitney) entre la altura de las barras. **C** Trayectorias utilizadas por los robots de 50 simulaciones disintas para tres niveles de DA tónica. Los asteríscos de colores indican las posiciones iniciales de cada trayectoria.

Capítulo 5

CONCLUSIONES

Se presenta un modelo de los lazos cortico-ganglios basales que considera la presencia de dos lazos simultáneos, con aprendizaje en pesos cortico-estriados, considerando efectos del nivel de dopamina tónica sobre las neuronas del estriado, núcleo de entrada de información hacia los ganglios basales. Los efectos considerados incluyen tanto variaciones en las conexiones sinápticas cortico-estriadas como en características asociadas a la función de transferencia de las neuronas del estriado.

Los resultados obtenidos en simulaciones de tareas de selección y aprendizaje muestran que variaciones en los niveles de dopamina tónica no afectan los procesos internos de aprendizaje, sino más bien modifican el cómo se utiliza la información aprendida. Estos efectos son asociables a variaciones en la razón entre exploración y explotación.

Mediante la simulación de una tarea de supervivencia de dos recursos, se comprueba además que el modelo propuesto es capaz de funcionar como un mecanismo de selección de acciones.

El controlador propuesto, al aplicar el modelo de lazos cortico-ganglios basales como mecanismo de selección, permite ajustar los niveles de dopamina internos del robot, permitiendo simular efectos asociados con trastornos de ansiedad.

Para la tarea de supervivencia considerada, y la estructura del controlador propuesto, la razón entre exploración y explotación asociada al comportamiento del robot regulada por el nivel de dopamina modula la esperanza de vida. Los datos adquiridos sugieren que existe un nivel específico de dopamina tal que, considerando un nivel constante, se maximiza la esperanza de vida.

5.1. Trabajo futuro

Con respecto al modelo de lazos cortico-ganglios basales, extender el modelo presentado para considerar un control automático de los niveles de dopamina tónica permitiría determinar posibles efectos y patrones de comportamiento en los niveles de actividad sináptica sobre cada una de las estructuras presentadas (corticales y subcorticales).

Además, controlar el nivel de dopamina le permitiría a un agente autónomo definir en forma dinámica, por ejemplo, la razón de exploración-explotación. Variaciones dinámicas de esta proporción permitirían favorecer la exploración en ambientes o periodos de tiempo que así lo requieran, como situaciones de alta incertidumbre. Del mismo modo, se podría favorecer la explotación en circunstancias donde se haya determinado que la explotación

aumenta los beneficios futuros. Considerando el alto desempeño de animales y humanos con respecto a la toma de decisiones en ambientes inciertos [9], la implementación de un control de dopamina bio-inspirado (e.g., [48]) puede ser una contribución en Inteligencia Artificial.

Apéndice A

APÉNDICE

A.1. Parámetros utilizados

Los parámetros utilizados en las simulaciones, durante la tarea de selección forzada aplicada en el capítulo 3, y la tarea de supervivencia de dos recursos aplicada en el capítulo 4, son presentados en la tabla A.1. Del conjunto de parámetros presentados en la tabla, las constantes de tiempo τ y los pasos de tiempo dt se encuentran medidas en segundos, n es el número de *alternativas*, y los umbrales T, la amplitud A y los parámetros de nivel de las funciones sigmoides (V) se encuentran en espigas por segundo $[sp/s]$. El resto de los parámetros son cantidades adimensionales.

A.2. Valores iniciales

Para el modelo de las interacciones cortico-ganglios basales, todas las actividades sinápticas son iniciadas en cero. Los pesos cortico-estriados iniciales se determinan de forma pseudo-aleatoria, siguiendo una distribución normal según $w \sim \mathcal{N}(0.5, 0.005)$.

Tabla A.1: **Parámetros**.

dt	0.001	$G^{Th_{cog}}_{Ctx_{cog}}$	0.4	λ_{Ctx}	0.01
dt_{VREP}	0.05	$G^{Th_{mot}}_{Ctx_{mot}}$	0.4	λ_{Str}	0.01
$\tau_{\eta_{Str}}$	0.03	$G^{GPi_{cog}}_{STN_{cog}}$	1.0	λ_{STN}	0.01
τ	0.01	$G^{GPi_{mot}}_{STN_{mot}}$	1.0	λ_{GPi}	0.03
w_{min}	0.25	$G^{Th_{cog}}_{GPi_{cog}}$	-0. 5	λ_{Th}	0.01
w_{max}	0.75	$G^{Th_{mot}}_{GPi_{mot}}$	-0. 5	$\lambda_{Str^{cog}_{mot}}$	40.0
$G^{Str_{cog}}_{Ctx_{cog}}$	1.0	$G^{Ctx_{cog}}_{Th_{cog}}$	$(4/n)\cdot 1.0$	T_{Ctx}	-3.0
$G^{Str_{mot}}_{Ctx_{mot}}$	1.0	$G^{Ctx_{mot}}_{Th_{mot}}$	$(4/n)\cdot 1.0$	T_{STN}	-10.0
$G^{Str_{ass}}_{Ctx_{ass}}$	1.0	V_{min}	0.0	T_{GPi}	10.0
$G^{Str_{ass}}_{Ctx_{cog}}$	0.2	V_{max}	20.0	T_{Th}	-40.0
$G^{Str_{ass}}_{Ctx_{mot}}$	0.2	$V_{c_{DA}}$	3.0	α_{LTD}	0.0006
$G^{STN_{cog}}_{Ctx_{cog}}$	1.0	$V_{c_{Ass}}$	3.0	α_{LTP}	0.0006
$G^{STN_{mot}}_{Ctx_{mot}}$	1.0	$V_{h_{DA}}$	21.0	α_c	0.05
$G^{GPi_{cog}}_{Str_{cog}}$	-2.0	$V_{h_{Ass}}$	18.5	α_{VE}	0.0005
$G^{GPi_{mot}}_{Str_{mot}}$	-2.0	$\check{V}_h$	18.5	α_{PE}	0.01
$G^{GPi_{cog}}_{Str_{ass}}$	-2.0	VE_0	0.8	A	16.0
$G^{GPi_{mot}}_{Str_{ass}}$	-2.0	PE_0	0.2	n	{4,5}

REFERENCIAS

[1] Hyungil Ahn and Rosalind W Picard. *Affective cognitive learning and decision making: The role of emotions*. na, 2006.

[2] Garrett E Alexander, Michael D Crutcher, and Mahlon R DeLong. Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor,"prefrontal" and "limbic" functions. *Progress in brain research*, 85:119–146, 1991.

[3] Garrett E Alexander, Mahlon R DeLong, and Peter L Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, 9(1):357–381, 1986.

[4] A Blanchard and Lola Canamero. Developing affect-modulated behaviors: Stability, exploration, exploitation or imitation. In *Procs 6th Int Workshop on Epigenetic Robotics*. Lund University, 2006.

[5] Rufino Bolado-Gomez and Kevin Gurney. A biologically plausible embodied model of action discovery. *Frontiers in neurorobotics*, 7, 2013.

[6] Paolo Calabresi, Diego Centonze, and Giorgio Bernardi. Electrophysiology of dopamine in normal and denervated striatal neurons. *Trends in neurosciences*, 23:S57–S63, 2000.

[7] Heng-Da Cheng, XH Jiang, Ying Sun, and Jingli Wang. Color image segmentation: advances and prospects. *Pattern recognition*, 34(12):2259–2281, 2001.

[8] Alexandre Coninx, Agnés Guillot, Benoît Girard, et al. Adaptive motivation in a biomimetic action selection mechanism. In *Deuxième conférence française de Neurosciences Computationnelles,'Neurocomp08'*, 2008.

[9] Peter Dayan and Nathaniel D Daw. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):429–453, 2008.

[10] M. Freese E. Rohmer, S. P. N. Singh. V-rep: a versatile and scalable robot simulation framework. In *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[11] Chris Eliasmith, Terrence C Stewart, Xuan Choo, Trevor Bekolay, Travis DeWolf, Yichuan Tang, and Daniel Rasmussen. A large-scale model of the functioning brain. *science*, 338(6111):1202–1205, 2012.

[12] Elodie Fino and Laurent Venance. Spike-timing dependent plasticity in the striatum. *Spike-timing dependent plasticity*, page 219, 2010.

[13] Vincenzo G Fiore, Valerio Sperati, Francesco Mannella, Marco Mirolli, Kevin Gurney, Karl Friston, Raymond J Dolan, and Gianluca Baldassarre. Keep focussing: striatal dopamine multiple functions resolved in a single mechanism tested in a simulated humanoid robot. *Frontiers in Psychology*, 5:124, 2014.

[14] Michael J Frank. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of cognitive neuroscience*, 17(1):51–72, 2005.

[15] Benoît Girard, Vincent Cuzin, Agnès Guillot, Kevin N Gurney, and Tony J Prescott. A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of integrative neuroscience*, 2(02):179–200, 2003.

[16] Benoît Girard, Nicolas Tabareau, Quang-Cuong Pham, Alain Berthoz, and J-J Slotine. Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection. *Neural Networks*, 21(4):628–641, 2008.

[17] F Montes Gonzalez, Tony J Prescott, Kevin Gurney, Mark Humphries, and Peter Redgrave. An embodied model of action selection mechanisms in the vertebrate brain. *From animals to animats*, 6:157–166, 2000.

[18] Kevin Gurney, Tony J Prescott, and Peter Redgrave. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, 84(6):401–410, 2001.

[19] Kevin Gurney, Tony J Prescott, and Peter Redgrave. A computational model of action selection in the basal ganglia. ii. analysis and simulation of behaviour. *Biological cybernetics*, 84(6):411–423, 2001.

[20] Martin Guthrie, Arthur Leblois, André Garenne, and Thomas Boraud. Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. *Journal of neurophysiology*, 109(12):3025–3040, 2013.

[21] James C Houk, Joel L Davis, and David G Beiser. *Models of information processing in the basal ganglia*. MIT press, 1995.

[22] Mark D Humphries, Mehdi Khamassi, and Kevin Gurney. Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in neuroscience*, 6, 2012.

[23] Shin Ishii, Wako Yoshida, and Junichiro Yoshimoto. Control of exploitation–exploration meta-parameter in reinforcement learning. *Neural networks*, 15(4):665–687, 2002.

[24] Yohan J John, Daniel Bullock, Basilis Zikopoulos, and Helen Barbas. Anatomy and computational modeling of networks underlying cognitive-emotional interaction. *Frontiers in human neuroscience*, 7, 2013.

[25] Roberta M Kelly and Peter L Strick. Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits. *Progress in brain research*, 143:447–459, 2004.

[26] Mehdi Khamassi, Stéphane Lallée, Pierre Enel, Emmanuel Procyk, and Peter F Dominey. Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Frontiers in neurorobotics*, 5, 2011.

[27] Jeffrey L Krichmar. A neurorobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior. *Frontiers in neurorobotics*, 7, 2013.

[28] Jeffrey L Krichmar and Florian Röhrbein. Value and reward based learning in neurorobots. *Frontiers in neurorobotics*, 7, 2013.

[29] Arthur Leblois, Thomas Boraud, Wassilios Meissner, Hagai Bergman, and David Hansel. Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. *The Journal of neuroscience*, 26(13):3567–3583, 2006.

[30] Henry Markram. A brain in a supercomputer, 2009.

[31] Jean-Arcady Meyer, Agnès Guillot, Benoît Girard, Mehdi Khamassi, Patrick Pirim, and Alain Berthoz. The psikharpax project: Towards building an artificial rat. *Robotics and autonomous systems*, 50(4):211–223, 2005.

[32] François Michaud, Paolo Pirjanian, Jonathan Audet, and Dominic Létourneau. Artificial emotion and social robotics. In *Distributed autonomous robotic systems 4*, pages 121–130. Springer, 2000.

[33] Jonathan W Mink. The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, 50(4):381–425, 1996.

[34] F Montes-Gonzalez, TJ Prescott, and J Negrete-Martinez. Minimizing human intervention in the development of basal ganglia-inspired robot control. *Applied Bionics and Biomechanics*, 4(3):101–109, 2007.

[35] Atsushi Nambu, Hironobu Tokuno, Ikuma Hamada, Hitoshi Kita, Michiko Imanishi, Toshikazu Akazawa, Yoko Ikeuchi, and Naomi Hasegawa. Excitatory cortical inputs to pallidal neurons via the subthalamic nucleus in the monkey. *Journal of neurophysiology*, 84(1):289–300, 2000.

[36] Saleem M Nicola, F Woodward Hopf, and Gregory O Hjelmstad. Contrast enhancement: a physiological effect of striatal dopamine? *Cell and tissue research*, 318(1):93–106, 2004.

[37] Eric S Nisenbaum and Charles J Wilson. Potassium currents responsible for inward and outward rectification in rat neostriatal spiny projection neurons. *The Journal of neuroscience*, 15(6):4449–4463, 1995.

[38] John P O'Doherty, Peter Dayan, Karl Friston, Hugo Critchley, and Raymond J Dolan. Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2):329–337, 2003.

[39] Patricio O'Donnell. Dopamine gating of forebrain neural ensembles. *European Journal of Neuroscience*, 17(3):429–435, 2003.

[40] Karla Parussel and Lola Cañamero. Biasing neural networks towards exploration or exploitation using neuromodulation. In *Artificial Neural Networks–ICANN 2007*, pages 889–898. Springer, 2007.

[41] Benjamin Pasquereau, Agnes Nadjar, David Arkadir, Erwan Bezard, Michel Goillandeau, Bernard Bioulac, Christian Eric Gross, and Thomas Boraud. Shaping of motor responses by incentive values through the basal ganglia. *The Journal of neuroscience*, 27(5):1176–1183, 2007.

[42] Tony J Prescott, Fernando M Montes González, Kevin Gurney, Mark D Humphries, and Peter Redgrave. A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Networks*, 19(1):31–61, 2006.

[43] Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S Mark Williams. *Neuroscience*. Sunderland, MA: Sinauer Associates, third edition, 2004.

[44] Fabian Rubilar, María-josé Escobar, and Tomás Arredondo. Bio-inspired architecture for a reactive-deliberative robot controller. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 2027–2035. IEEE, 2014.

[45] Michael I Sandstrom and George V Rebec. Characterization of striatal activity in conscious rats: Contribution of nmda and ampa/kainate receptors to both spontaneous and glutamate-driven firing. *Synapse*, 47(2):91–100, 2003.

[46] Henning Schroll and Fred H Hamker. Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy. *Frontiers in systems neuroscience*, 7, 2013.

[47] Wolfram Schultz. Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, 80(1):1–27, 1998.

[48] Nicolas Schweighofer and Kenji Doya. Meta-learning in reinforcement learning. *Neural Networks*, 16(1):5–9, 2003.

[49] Jesus Soares, Michele A Kliem, Ranjita Betarbet, J Timothy Greenamyre, Bryan Yamamoto, and Thomas Wichmann. Role of external pallidal segment in primate parkinsonism: comparison of the effects of 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-induced parkinsonism and lesions of the external pallidal segment. *The Journal of neuroscience*, 24(29):6417–6426, 2004.

[50] Emmet Spier and DJ McFarland. A finer-grained motivational model of behaviour sequencing. In *Proceedings of fourth International Conference on Simulation of Adaptive Behavior*, pages 255–263, 1996.

[51] Larry Squire, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. *Fundamental neuroscience*. Academic Press, third edition, 2008.

[52] Terrence C Stewart, Trevor Bekolay, and Chris Eliasmith. Learning to select actions with spiking neurons in the basal ganglia. *Frontiers in neuroscience*, 6, 2012.

[53] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[54] Robert S Turner and Marjorie E Anderson. Pallidal discharge related to the kinematics of reaching movements in two dimensions. *Journal of neurophysiology*, 77(3):1051–1074, 1997.

[55] Toby Tyrrell. The use of hierarchies for action selection. *Adaptive Behavior*, 1(4):387–420, 1993.

[56] Yiping Wang, Sheng Li, Qingwei Chen, and Weili Hu. Biology inspired robot behavior selection mechanism: using genetic algorithm. In *Bio-Inspired Computational Intelligence and Applications*, pages 777–786. Springer, 2007.

[57] Marc G Weisskopf, Honglei Chen, Michael A Schwarzschild, Ichiro Kawachi, and Alberto Ascherio. Prospective study of phobic anxiety and risk of parkinson's disease. *Movement disorders*, 18(6):646–651, 2003.

[58] Charles J Wilson and Philip M Groves. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain research*, 220(1):67–80, 1981.

[59] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1, 1972.