Comportamiento emergente de acercamiento a un objeto en un modelo de pez eléctrico mediante aprendizaje por refuerzo

Ignacio Naya

Orientador: Dr. Leonel Gómez-Sena

Coorientador: Ing. Gonzalo Tejera

Tesina de finalización de carrera - Licenciatura en Ciencias Biológicas

Profundización en Biomatemática

Facultad de Ciencias, Universidad de la República

Montevideo, Uruguay

Diciembre, 2019

Agradecimientos

A mi orientador, Leonel Gómez,

a mi coorientador, Gonzalo Tejera,

a los miembros del tribunal, Adriana Migliaro y Fernando Álvarez,

a mi madre, padre, hermanos y demás familiares,

a aquellos otros por los que uno ha adquirido y va adquiriendo afecto

y a las circunstancias que lo permiten.

Tabla de contenidos

1. Resumen **6**

2. Introducción **8**

2.1. El Sistema Nervioso y la movilidad en animales 8

2.2. Recepción sensorial 9

2.3. Electrolocalización 10

2.4. Gnathonemus petersii 11

2.4.1. Generalidades y electrorreceptores 11

2.4.2. Electrorreceptores Mormyromast y descargas del órgano eléctrico en Gnathonemus petersii (Mormiridae) 14

2.4.3. Patrones motores 17

2.5. Toymodel 18

2.6. Problema a abordar 20

2.7. Aprendizaje por refuerzo 23

2.7.1. Aprendizaje Q (Q-learning) 24

2.7.2. Red neuronal artificial de base radial 27

3. Objetivos generales y objetivos específicos **33**

3.1. Objetivos generales 33

3.2. Objetivos específicos 33

4. Materiales y Métodos **34**

4.1. Configuración del agente y su arena 34

4.2. Parámetros 37

4.3. Evaluación rendimiento a corto y mediano plazo 38

4.4. Alineamiento medio en relación a la distancia 38

4.5. Ángulo de giro (Φ) en relación al ángulo de alineación (θ) 39

5. Resultados **41**

5.1. Observaciones cualitativas del comportamiento del agente 41

5.2 Rendimiento a corto y mediano plazo 43

5.3. Parámetros 45

5.4. Alineamiento medio en relación a la distancia 53

5.5. Ángulo de giro (Φ) en relación al ángulo de alineación (θ) 56

6. Discusión **58**

6.1. Parámetros 59

6.2. Rendimiento del algoritmo 60

6.3. Alineación en relación con la distancia 61

6.4. Ángulo de giro (Φ) en relación al ángulo de alineación (θ) 63

6.5. Toymodel 65
6.6. Aplicación del modelo en relación a aspectos biológicos 66

7. Conclusiones 70

8. Perspectivas 71

9. Bibliografía 71

10. Anexo 76

1. Resumen

Una de las características que distingue a los vertebrados es su desarrollada capacidad de locomoción que mejora su aptitud. Esta es dependiente del sistema nervioso que para lograr navegar en un determinado entorno debe conseguir hacer en general tres cosas: recibir señales del medio (recepción de estímulos), procesarlas y, a partir de este procesamiento, enviar señales a las unidades motoras (efectoras). Esta recepción puede clasificarse como pasiva cuando no hay emisión de energía o activa cuando los animales “sensan” el ambiente mediante la emisión de energía que actúa como portadora de los estímulos. Un caso particular de esto último es la electrolocalización en los peces eléctricos de descarga débil. Esta se basa en la emisión de descargas eléctricas de baja amplitud que son posteriormente moduladas por el ambiente generando una distribución de corrientes eléctricas sobre la superficie transcutánea de estos peces que es captada por los electrorreceptores. A esta distribución de corrientes se le llama imagen eléctrica y permite obtener la ubicación espacial, así como distintas propiedades físicas de aquellos objetos que se encuentran en torno a estos. Una de las especies pertenecientes a la categoría de peces eléctricos, que inspiró el presente estudio, es *Gnathonemus petersii*, que posiblemente, de manera de reducir ambigüedades y por ende obtener información acerca de tales objetos, presentan distintos patrones comportamentales que refinan el proceso perceptivo. El presente estudio se enfoca en particular en el patrón de acercamiento a un objeto. Más precisamente en dos de sus características: el acercamiento y el alineamiento dependiente de la distancia. Tratando especialmente dos propiedades que se sospecharon en principio relacionadas, la primera consistía en mover el pico de la imagen eléctrica hacia la posición de la cabeza a medida que el pez avanza hacia el objeto (lo que se relaciona con un alineamiento con el objeto) y la segunda en aumentar la amplitud de la imagen eléctrica total. Paralelamente en otro modelo implementado anteriormente, que ajustaba su navegación aplicando una función sigmoidal dependiendo del alineamiento que este presentaba respecto a un determinado objeto, se había logrado recrear el acercamiento y el alineamiento dependiente de la distancia. En este estudio se

consideran ambos problemas en conjunto: el hecho de mover el pico de la imagen eléctrica hacia la posición de la cabeza como una propiedad emergente del aumento de la amplitud de la imagen eléctrica y el hecho de obtener una función de corrección de error relativamente similar a la utilizada por Gómez-Sena (similar a una función sigmoide) mediante algún algoritmo de aprendizaje. Ambos problemas se abordan mediante el modelado de un pez con características similares al modelo ya descrito, que recibe determinada premiación al aumentar la amplitud de la imagen eléctrica y que implemente un algoritmo de aprendizaje que aplica una red neuronal con un filtro de base radial. Tratando de responder si se puede hacer emerger el comportamiento de tal modelo, analizando para ello si se registran movimientos de rotación dependiente del alineamiento y si el hecho de premiar el aumento de la imagen eléctrica es suficiente como para explicar el alineamiento dependiente de la distancia, haciéndose también un relevamiento del rendimiento del programa (medido en recompensas obtenidas por iteración realizada) para los distintos parámetros.

Con respecto a los movimientos de rotación dependiente del ángulo se logra obtener al analizar un conjunto de iteraciones, en promedio y en general, una rotación en el sentido esperado. No siendo este el caso para ángulos en los que el “pez” se encontraba “bastante” alineado al objeto, en donde no se registró un ángulo de rotación tendiente a alinearlo al mismo. Aparte, no pudo registrarse un alineamiento notorio respecto al acercamiento del pez al objeto. Por tanto, no se puede afirmar que el hecho de premiar el aumento de la imagen eléctrica sea suficiente para explicar la emergencia del alineamiento dependiente de la distancia. Aunque sí, mediante el algoritmo empleado, se logra una corrección del sentido de rotación dependiente del alineamiento.

2. Introducción

2.1. El Sistema Nervioso y la movilidad en animales

Una de las características que distingue a los vertebrados es su desarrollada capacidad de locomoción. Esta capacidad permite a éstos una mayor probabilidad de alcanzar los recursos y condiciones que aumentan su aptitud (Zug, 2017). En conjunto con esta capacidad de movimiento (especialmente la navegación), se ha desarrollado a lo largo de la evolución, probablemente hace más de 550 millones de años, un sistema de células conectadas entre sí, especializadas en la conducción de señales eléctricas: el sistema nervioso (Northcutt, 2012). Este sistema de células que en un principio apareció de forma reticular, tal como se aprecia en cnidarios y ctenóforos, fue adquiriendo a lo largo de la evolución una mayor centralización (Lentz and Erulkar, 2019). De esta forma, posteriormente a la aparición de este sistema, surge la cefalización (característica de los animales bilaterales), caracterizada, en el esquema más simple (cilindro con parte anterior donde se posiciona la boca y posterior donde se posiciona el ano), por presentarse en la parte anterior de los animales con simetría bilateral (Lentz and Erulkar, 2019).

La cefalización está en cierta medida relacionada con la capacidad de desplazamiento en los animales (de nuevo, especialmente la navegación), ya que si bien en cnidarios está presente dicha capacidad, es en los animales con distintos grados de cefalización donde se observa mayor desarrollo de aspectos más finos relativos a la navegación (Breed and Moore, 2016). Además, como ejemplo de esta relación (cefalización y navegación) se puede apreciar el estrecho vínculo entre la cefalización y la navegación en tunicados (Hofmann et al., 2013), cuyo sistema nervioso se ve altamente reducido luego de fijarse a un sustrato en los últimos estadíos de la fase larval del

mismo (MacIver, 2008). También aporta en esta línea argumental observar que los moluscos sésiles presentan un menor desarrollo del sistema nervioso que aquellos pelágicos (Breed & Moore, 2016).

Wyse et al. (2013) describe de manera concisa las características funcionales del sistema nervioso de la siguiente forma:

- Las neuronas están organizadas en circuitos de tal manera que ellas pueden provocar una respuesta coordinada y adaptativa de los efectores.
- Las células receptoras sensoriales transforman los estímulos del ambiente en señales eléctricas.
- Las interneuronas centrales integran las señales provenientes, tanto de los receptores sensoriales como de otras señales que se originan dentro del animal, generando así un patrón de impulsos integrados.
- Los comandos motores son enviados fuera del sistema nervioso central hacia los efectores

Por tanto, el sistema nervioso para lograr navegar en un determinado entorno debe lograr hacer en general tres cosas: recibir señales del medio, procesarlas y, a partir de este procesamiento, enviar señales a las unidades motoras (efectoras), las cuales permitirán al organismo trasladarse siguiendo algún procedimiento efectivo, que mejore la adaptabilidad del individuo en relación al ambiente (Bowdan & Wyse, 1996; Kandel et al., 2013).

2.2. Recepción sensorial

Los sentidos y los receptores sensoriales, son los medios por los cuales los animales detectan y responden a los estímulos en sus entornos internos y externos, por lo que son necesarios para efectuar tales tareas de navegación (Land, 2019). La recepción sensorial de estímulos, lo que

sería el “sensado” (del inglés, sensing), es la consecuente recepción efectuada por tales medios. Esta puede clasificarse de acuerdo a la actividad del organismo a la hora de interactuar con el ambiente (Hofmann et al., 2013). En concordancia con Hofmann et al., (2013), el “sensado” del entorno que rodea a los individuos se puede dividir básicamente en dos tipos: uno de ellos es el “sensado” pasivo, el cual incluye a todos aquellos tipos en los que los estímulos son modulaciones de un patrón de energía generada por una fuente externa y un segundo tipo correspondiente al “sensado” activo, en donde la energía portadora de los estímulos es generada por el propio organismo. En este último caso el proceso de generación de la señal que terminará siendo captada por el organismo, consiste en la emisión de energía en una determinada forma que sirve como portadora de una señal en cierto estado basal (determinado por el organismo) que posteriormente será modulada por el ambiente. Ejemplos de este tipo de señal son la ecolocalización por parte de los murciélagos, el tanteo que los insectos realizan con sus antenas, el sentido del tacto exploratorio y la electrolocalización activa por parte de peces eléctricos de las familias Gymnotidae y Mormyridae (Wyse et al., 2013; Hofmann et al., 2013).

2.3. Electrolocalización

Como indican Gómez-Sena et al., (2014): Los organismos han evolucionado su capacidad de obtener información del ambiente “sensando” cambios en los distintos tipos de energía.

Un caso particular es la capacidad que exhiben algunas especies de peces de detectar cambios en el campo eléctrico generado por otros animales los animales cuando sus tejidos muscular y nervioso generan potenciales eléctricos. Estos pueden ser detectados en un medio conductivo como lo es el agua. Esta capacidad de detectar campos eléctricos emitidos por agentes externos y localizar su fuente, sin emisión de energía por parte del pez en cuestión, entra dentro de la categoría de “sensado” pasivo y es la llamada electro-localización pasiva (Hopkins, 2005; Hofmann et al., 2013).

No obstante, existe además un grupo de peces que, no solo son capaces de detectar objetos de forma pasiva, sino que además son capaces de emitir energía, en forma de campos eléctricos (Hofmann et al., 2013). Estos últimos son los llamados peces eléctricos que están constituidos (considerando solo a los pertenecientes a la clase Actinopterygii) principalmente por las familias Gymnotidae (proveniente de América), Mormyridae (proveniente de África) y Gymnarchidae (también de África), estas dos últimas pertenecientes a la superfamilia Mormyroidea. Los peces de dicha superfamilia han adquirido esta capacidad de forma independiente de aquellos pertenecientes a la Gymnotidae, aunque posiblemente ambos lo hicieron a partir de electrorreceptores pasivos (Lissmann, 1951; Lissmann, 1958).

2.4. *Gnathonemus petersii*

2.4.1. Generalidades y electrorreceptores

Gnathonemus petersii es una especie pez, de origen africano de la familia Mormyridae, el cual ha sido uno de los principales modelos de estudio acerca de peces eléctricos. Esta especie de hábitos nocturnos se encuentra en aguas que están caracterizadas por presentar una baja conductancia, suficiente para establecer un campo eléctrico que le es útil, como a otros peces eléctricos, para distintas funciones que le sirven para adaptarse al medio (Lissmann, 1951; Engelmann & von der Emde, 2011; Hofmann et al., 2013).

Estos peces entran dentro de la categoría de los peces eléctricos de descarga débil, dado que la amplitud de la descarga eléctrica es relativamente baja en comparación con aquellos peces considerados de descarga eléctrica fuerte (Zupanc & Bullock, 2005). Están dotados de numerosos electrorreceptores distribuidos sobre la cabeza y el cuerpo (figura 1) en diferentes densidades (Bacelo et al., 2008). Dichos receptores pueden dividirse en tres tipos de electrorreceptores anatómica y funcionalmente separados, relacionados con la actividad eléctrica (Bell, 1989):

- Los receptores de tipo ampular
- Los electrorreceptores del órgano de Knollen (“Knollenorgan electroreceptors”)
- Los electrorreceptores Mormyromast (“Mormyromast elctroreceptrors”)

El primero de los tres órganos está presente en una amplia cantidad de peces, y este no requiere de emisión de descargas eléctricas para funcionar, es decir, el tipo de “sensado” es pasivo, como ya se indicó en la sección anterior (Bell, 1989). El órgano de Knollen se considera que tiene funciones relacionadas principalmente con la comunicación ya que, entre otras cosas, tiene la capacidad de codificar eficientemente la información temporal de los estímulos eléctricos. No obstante, la intensidad del estímulo parece ser ampliamente ignorada por este electrorreceptor así como también la información espacial (Bell, 1989). El tercero de los electrorreceptores está relacionado con la electrolocalización y se profundizará en la sección siguiente.

Cabe destacar que los tres electrorreceptores actúan junto con el órgano eléctrico, el cual genera descargas eléctricas (generando ruido para el primer tipo de estos receptores, que el pez debe corregir). Esto último determina además que el “sensado” realizado por el último de los electrorreceptores sea activo (ya que se basan en su descarga para obtener información del medio).

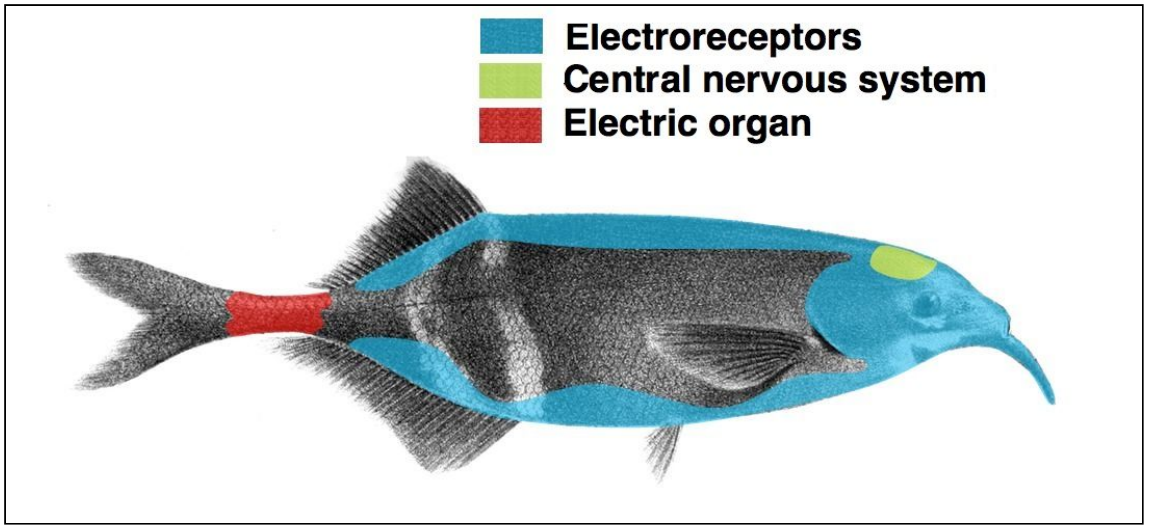

Figura 1. Morfología del pez *Gnathonemus Petersii* y ubicación anatómica electrorreceptores (cian), sistema nervioso central (verde) y órgano eléctrico (rojo). Extraído de https://en.wikipedia.org/wiki/Peters%27_elephantnose_fish#/media/File:Gnathonemus_petersii_(G%C3%BCnther,_1862).en2.jpg

2.4.2. Electrorreceptores Mormyromast y descargas del órgano eléctrico en *Gnathonemus petersii* (Mormiridae)

A partir de las controladas, breves emisiones bifásicas, de pulsos eléctricos generadas por el órgano eléctrico, se generan campos eléctricos en torno al pez (Post & von der Emde, 1999; Rother et al., 2003; Engelmann et al., 2016). Estos campos se ven modulados, por los objetos cercanos al pez (debido a que presentan distintas inductancias, formas, etc) en relación a la condición basal, resultando en una alteración local en la distribución de la corriente sobre la piel del pez (Gómez-Sena et al., 2004; Engelmann & von der Emde, 2011; Engelmann et al., 2016). Siendo lo relevante de esta distribución de corrientes transcutáneas consiste en que la misma se distribuye de forma bidimensional a lo largo y ancho de todo el arreglo sensorial del mismo (Gómez-Sena et al., 2004; Caputi & Budelli; 2006). A esta distribución de corrientes se le suele llamar imagen eléctrica y para objetos simples, concordando con Engelmann et al (2016), suele tener la forma de un gorro mexicano (Rasnow, 1996; Caputi et al., 1998; Budelli and Caputi, 2000; Rother et al., 2003; Gómez-Sena et al., 2004; Caputi & Budelli; 2006). La forma de gorro mexicano es debida a que en la zona central se suele generar un aumento en la densidad de líneas de corriente (figura 2), para materiales más conductores que el medio que rodea al pez, mientras que en la periferia ocurre lo contrario. Explicación similar se puede suponer para el caso contrario, en donde el objeto reduce el paso de corriente en la zona central, generando un pico negativo en la zona central y en la periferia se registra, por lo tanto, una mayor densidad de corriente (Caputi et al., 1998; von der Emde, 1999). La zona de mayor amplitud (en términos absolutos), correspondiente a la proyección del pico de la imagen eléctrica, indica la ubicación sin el componente de profundidad (Maler, 2009a, 2009b, Engelmann et al., 2016).

Para lograr determinar la profundidad (la distancia a la que se encuentra del objeto), lo que en principio se sugirió es que estos peces, así como otros peces de esta familia, utilizan la relación entre la pendiente máxima y la amplitud máxima (amplitud en sentido físico) de la imagen eléctrica (von der Emde, 1999). Siendo esta relación cada vez menor a distancias cada vez mayores,

teniendo un rango de detección aproximadamente igual o el doble al largo del pez (largo generalmente cercano a los 12cm) (von der Emde, 1999). Posteriormente, Hofmann et al., (2017) han propuesto otra forma (no excluyente) de estimación de la distancia hacia el objeto, en la que esta estaría determinada por el gradiente relativo y por ende también en parte por el flujo electrosensorial, teniendo una gran importancia el comportamiento del pez a la hora de darle forma a la información sensorial.

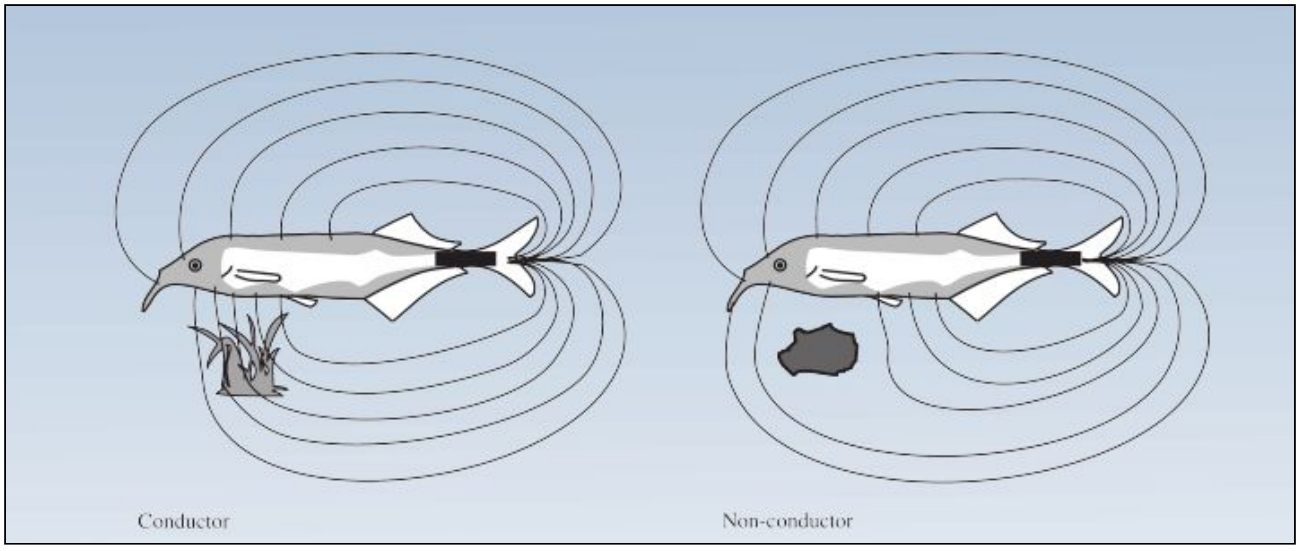

Figura 2. Alteración en el patrón de las líneas de campo eléctrico (líneas curvas) generado por el pez *Gnathonemus petersii* debido a un material conductor y a uno no conductor. Línea negra gruesa de la cola representa el órgano eléctrico y en gris claro se representa a los receptores eléctricos (von der Emde, 1999).

2.4.3. Patrones motores

El comportamiento de estos peces parece tener influencia a la hora de obtener información acerca de los objetos, siendo este proceso dependiente de la cinética del pez, presentando distintos patrones motores (Hofmann, 2013; Hofmann et al., 2014). Dichos patrones, analizados mediante métodos semicuantitativos y cualitativos (Toerring and Belbenoit, 1979; Toerring and Moller, 1984; von der Emde, 1992; Hofmann et al., 2014) se lograron definir en forma cuantitativa (cinemáticamente por un patrón de movimientos prototípicos) en el estudio realizado por Hofmann et al. (2014). En este último se identificaron, entre otros, los siguientes patrones comportamentales: nado en reversa, comportamiento estacionario y comportamiento de aproximación al objeto.

El nado en reversa consiste en el movimiento del pez con una velocidad de avance ("thrust") con módulo negativo. El comportamiento estacionario, en cambio, está formado por una muy baja (en términos absolutos) velocidad de avance ("thrust").

El último comportamiento (aproximación al objeto) se registra cuando estos peces se encuentran con objetos novedosos, es decir, el comportamiento está dirigido a un objeto (Hofmann et al., 2014) e implica cambios en el comportamiento de muestreo (Figura 4) que además dependen de la distancia al objeto en cuestión (Hofmann et al., 2017). Dichos cambios comportamentales incluyen la navegación hacia el objeto, un alineamiento con respecto al objeto mencionado de forma de dejar la región cefálica (en dónde se encuentra la fóvea electrosensorial y donde la canalización es máxima) más cercana al objeto y reduciendo la curvatura del pez, un aumento en la frecuencia de las descargas por parte del órgano eléctrico y una disminución de la velocidad de avance a medida que el pez se acerca a dicho objeto (Castelló et al., 2000; Caputi et al., 2002; Hofmann et al., 2017).

Partiendo de esto último, se ha sugerido como posible regla que este comportamiento está basado en las siguientes operaciones (Hofmann et al., 2017):

1. Mover el pico de la imagen eléctrica hacia la posición de la cabeza a medida que avanza hacia el objeto.
2. Aumentar la amplitud de la imagen eléctrica total.

2.5. Toymodel

De acuerdo con esto, Gómez-Sena (comunicación personal) simuló un pez (modelo denominado Toymodel) de forma circular (mediante el programa Octave) con sensores distribuidos por todo el perímetro de dicho círculo de forma que se moviera en un espacio bidimensional (Figura 5). Sus movimientos estaban mediados por un vector de módulo fijo cuyo ángulo de rotación se corregía comparando el ángulo (θ) que se forma entre el vector que parte del centro hasta la cabeza del "pez" (Figura 3) y el vector que conecta el centro del “pez” con el centro del objeto (ángulo que se correlaciona negativamente con el alineamiento del pez al objeto) utilizando para ello una función sigmoide como modo de corrección del ángulo con el que se efectúa la próxima acción, reduciendo de esta forma el ángulo θ y aumentando, por tanto, el alineamiento. Por lo que este modelo lograba crear un agente capaz de alinearse al objeto a medida que se acercaba al objeto mediante una función de corrección del ángulo de giro.

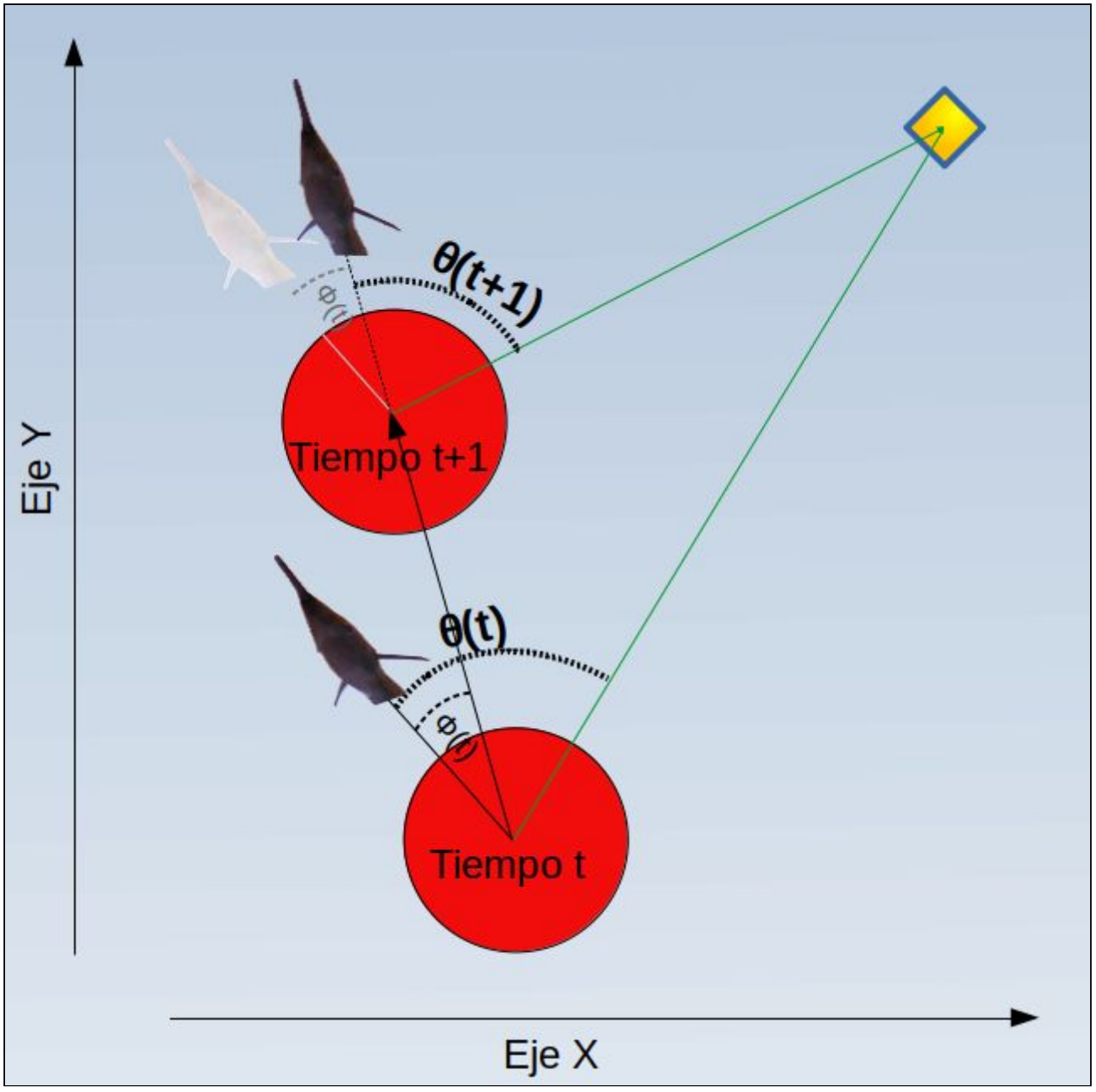

Figura 3. Descripción esquemática del ángulo theta. Visualización de cómo se obtiene la posición de la cabeza en el tiempo (t+1) luego de que el agente efectuase un desplazamiento en el tiempo (t). También se muestra el ángulo de rotación del vector desplazamiento respecto al vector que tiene su origen en el centro del "pez" y su extremo en el centro del objeto.

2.6. Problema a abordar

En este estudio se tratarán dos problemas simultáneamente: por un lado, se tratará de averiguar si el hecho de mover el pico de la imagen eléctrica hacia la posición de la cabeza (operación 1; descrita en la subsección 2.4.3) surge como una propiedad emergente de la operación 2 (descrita en la subsección 2.4.3), lo que se puede traducir como la emergencia de un alineamiento dependiente de la distancia a causa del aumento de la amplitud de la imagen eléctrica. En conjunto con esto, se estudiará la emergencia de una función de corrección de error relativamente similar a la utilizada por Gómez-Sena (similar a una función sigmoide) mediante la aplicación de algún algoritmo de aprendizaje. Tratando, en suma, de determinar si las condiciones mínimas necesarias para lograr emerger a un comportamiento razonablemente similar al comportamiento del “Toymodel” y, por lo tanto, a un comportamiento similar al observado de manera experimental, en un sistema capaz de ejercer un aprendizaje, consiste simplemente en premiar el aumento de la imagen.

Ambos problemas se abordarán mediante el modelado de un pez con características similares al “Toymodel”, el cual reciba determinada premiación al aumentar la amplitud de la imagen eléctrica y que implemente un algoritmo de aprendizaje por refuerzo. Planteándose específicamente el problema de la existencia de la emergencia del alineamiento dependiente de la distancia así como la corrección del sentido de giro dependiente del alineamiento.

Se sugiere además a esta regla como candidata sobre posibles mapas cognitivos ya que, bajo las condiciones experimentales modeladas por Hofmann et al. (2017), no existen otras referencias más que el objeto “objetivo” del pez y por lo tanto no se tendrían elementos para construir dicho mapa.

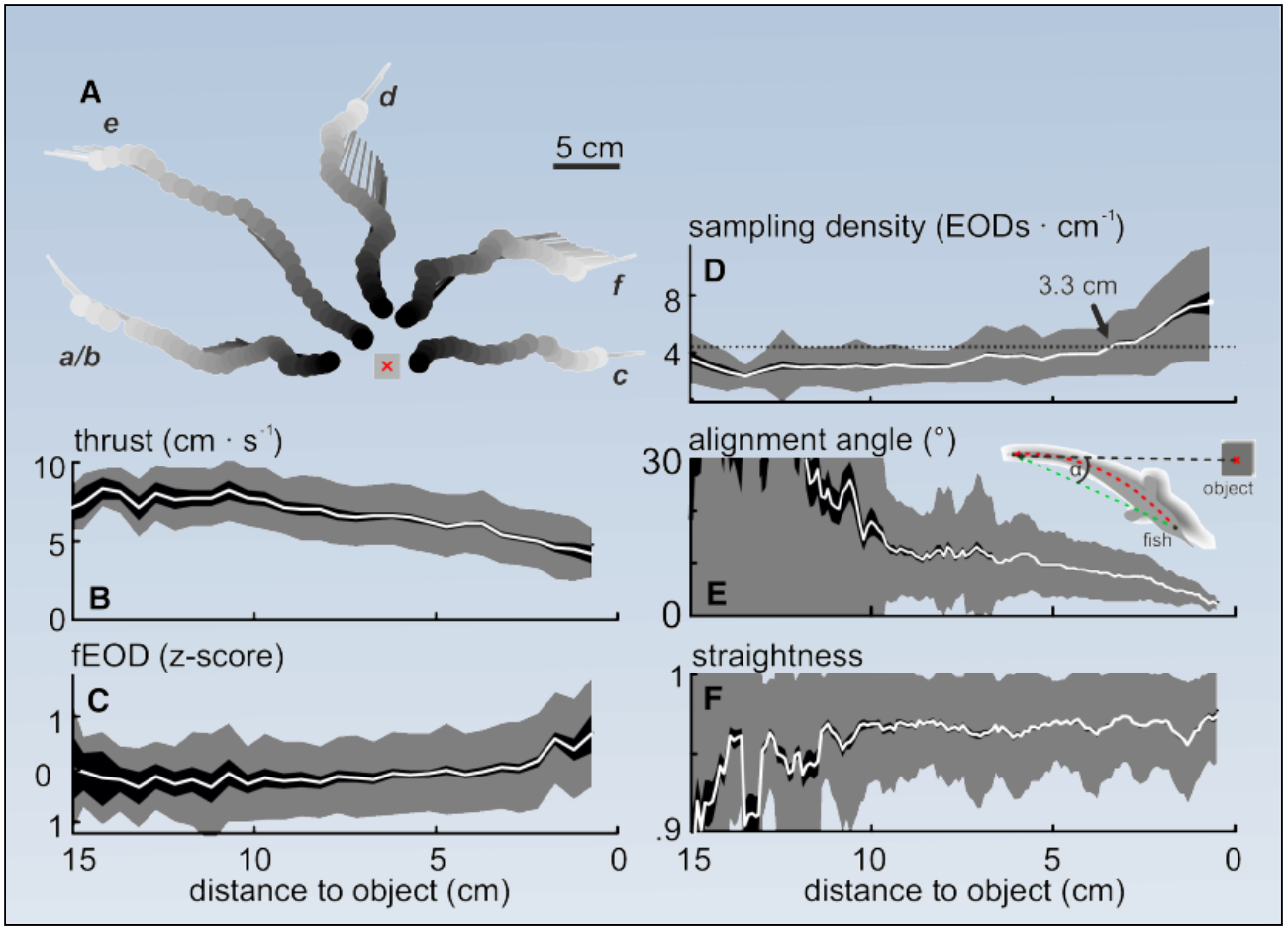

Figura 4. Gráfico adaptado del estudio efectuado por Hofmann et al. (2017): Comportamiento de acercamiento a un objeto. A) Distintos ejemplos de trayectorias del pez, los puntos indican la posición de la cabeza (la escala de grises representa la posición temporal, eventos más oscuros son los últimos en ocurrir), las líneas indican la orientación del cuerpo del pez y la cruz representa el centro de masa del objeto. B) Velocidad de avance del pez en relación a la distancia al objeto C) Frecuencia de muestreo efectuada por el pez con respecto a la distancia al objeto D) Densidad de muestreo (descargas eléctricas efectuadas en una determinada distancia recorrida) en relación a la distancia del pez al objeto E) Ángulo de alineación del pez con respecto al objeto, definido en este caso (distinto a como se define más abajo en el presente trabajo) como el ángulo formado entre la recta que se forma entre la cabeza y la cola del pez con la recta que une a la cola con el centro de masa del objeto F) Rectitud del pez, medido como el cociente entre la distancia de la cabeza a la cola sobre la longitud del pez, en relación a la distancia al objeto.

B-F) En blanco: Media; en negro: error estándar de la media; en gris: desvío estándar.

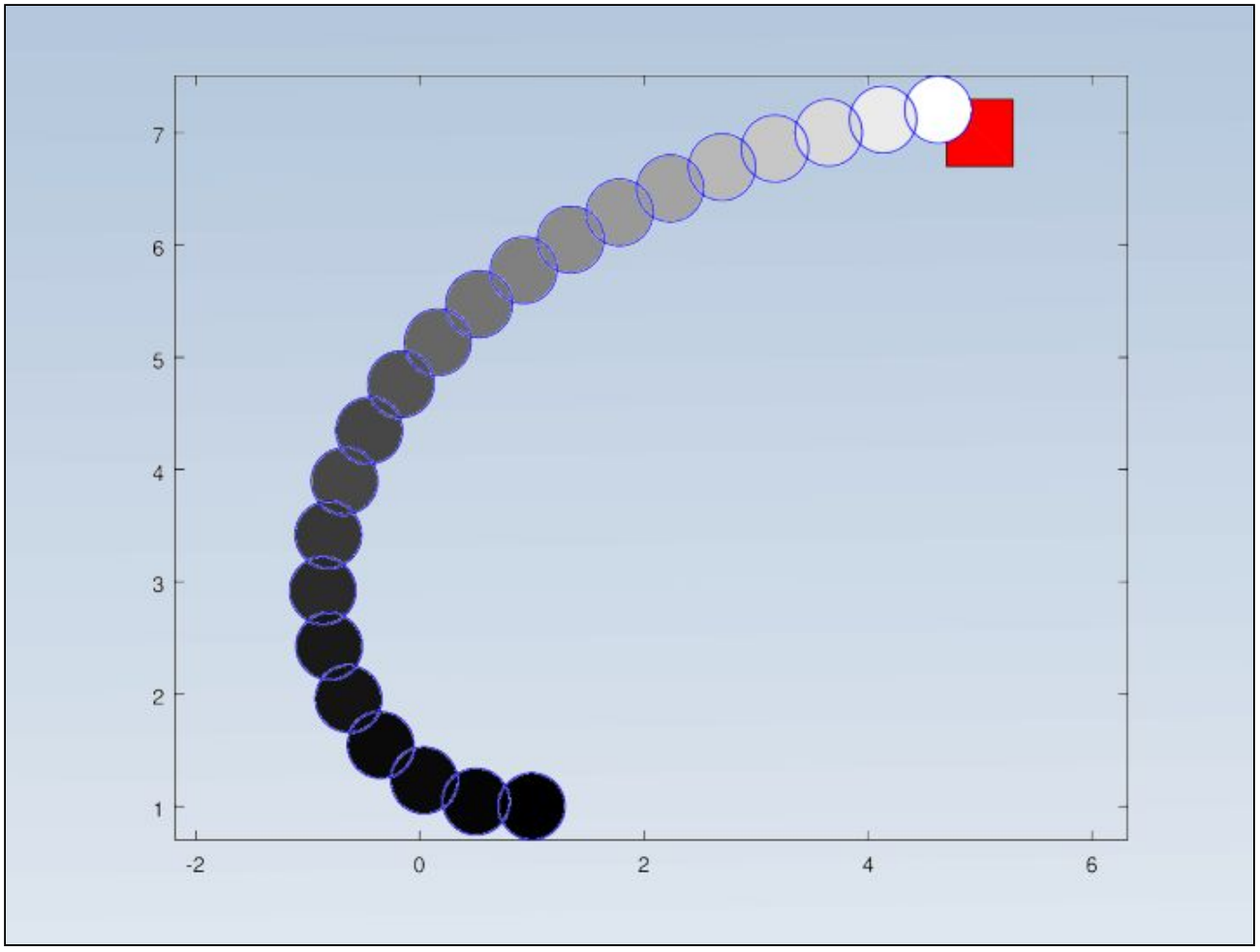

Figura 5. Ejemplo de una trayectoria obtenida a partir del programa realizado por Gómez-Sena, en escala de grises orden temporal de la trayectoria del pez (los valores más claros corresponden a las iteraciones más tardías temporalmente); los sensores no están visibles ni la orientación del pez (aunque esta última se puede inferir aproximadamente, al suponer que el pez se mueve siempre con la cabeza indicando el sentido del vector desplazamiento).

2.7. Aprendizaje por refuerzo

En la presente subsección se dará una introducción al aprendizaje por refuerzo en general, para así posteriormente irse adentrando en la lógica del algoritmo de aprendizaje que se utilizará en el presente trabajo.

Como indican Woergoetter y Porr (2008): Dado un determinado agente, el aprendizaje por refuerzo consiste en el aprendizaje realizado por éste mediado por la interacción con el ambiente que lo rodea. Aprendiendo las consecuencias de sus acciones, en vez de ser explícitamente enseñado y seleccionando sus acciones basándose en sus experiencias pasadas (explotación), así como a partir de la toma de nuevas decisiones (exploración), lo cual es esencialmente un aprendizaje de prueba y error. La señal de refuerzo que el agente recibe es un valor numérico que codifica el éxito del estado resultante de aplicar una determinada acción. Lo que el agente busca aprender es a seleccionar aquellas acciones que maximicen la recompensa acumulada a lo largo del tiempo.

Un caso particular de este tipo de aprendizaje es el algoritmo basado en refuerzos llamado "q-learning", que como otros, evalúa su política de acciones (el mapeo que realiza el agente de los distintos estados a distintas acciones) de forma continua, dándole mayor ponderación a la hora de actualizar su política a los estados (y acciones) que están más cercanos temporalmente. Las particularidades de dicho algoritmo serán explicadas a continuación, explicando para ello qué es la función de valor (así como el valor óptimo) y, obteniendo a partir de esta, la consiguiente función Q (y la función Q óptima), a partir de la cual se construye tal algoritmo (no obstante, si se quiere una mejor comprensión del mismo, se recomienda enfáticamente la lectura del capítulo "Aprendizaje por refuerzo" de Mitchell (1997)). Por otro, lado en la subsección 2.7.2 se explicará el algoritmo derivado de este, implementado por Santos y Touzet (1999b), que se empleará en el presente estudio.

2.7.1. Aprendizaje Q (Q-learning)

El aprendizaje por refuerzo trata especialmente el problema en el que la función de transición de un estado a otro (el estado que se obtiene al realizar determinada acción), así como la función de recompensa (la función que asigna un valor numérico, a cada acción que el agente toma en cada estado) son desconocidas para el agente, problemas denominados libres de modelos (Woergoetter y Porr, 2008). La función de valor para una política π en el estado s_t, $V^{\pi}(s_t)$, en el caso en que la función de transición y de recompensa sean deterministas (así como si nos encontramos ante un problema del tipo MDP[1]), suele ser definida como (Mitchell, 1997):

$$V^{\pi}(s_t) \equiv r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots \equiv \sum_{i=0}^{\infty} \gamma^i r_{t+i} \quad (1)$$

Siendo r_{t+i} la recompensa en el tiempo t+i (t e i números enteros no negativos), γ el factor de descuento (0<γ<1) y π una política dada. Es decir, la ecuación 1 describe la suma de recompensas obtenidas al aplicar una determinada política partiendo de cierto estado "s" en el tiempo t. Esta es una suma ponderada, que considera a las recompensas más cercanas temporalmente al estado actual como componentes más "relevantes" de dicha suma.

La política óptima (π*) en este contexto se define como:

$$\pi^* \equiv \underset{\pi}{argmax}\, V^{\pi}(s), (\forall s) \quad (2)$$

en donde π representa una política cualquiera y "s" un estado dado.

Por tanto (en concordancia con Mitchell (1997)), a partir de esto se tiene que el valor óptimo de $V^{\pi^*}(s)$ será aquél que se obtenga al aplicar la política óptima, es decir aquella política que maximice la suma de recompensas ponderadas.

---

[1] El proceso de decisión de Markov (MDP) es un proceso de control estocástico de tiempo discreto. Proporciona un marco matemático para modelar la toma de decisiones en situaciones donde los resultados pueden ser en parte aleatorios y están en parte bajo el control de un agente tomador de decisiones.

De esto se deduce que la política óptima π*: S→A (que selecciona a partir de cada estado determinada acción), en un MDP, en el estado s, se puede descomponer como:

$$\pi^* \equiv \underset{\pi}{argmax}\left[r(s,a)+\gamma V^*(\delta(s,a))\right] \tag{3}$$

en este caso se agrega la función r(s,a) que toma en consideración la recompensa de efectuar la acción "a", así como la función transición δ(s,a) que obtiene el valor del nuevo estado producto de aplicar dicha función al estado “s” junto con la acción "a".

Puede pensarse V como la función a ser aprendida para evaluar la transición entre estados. Dados dos estados posibles (s_1, s_2) si $V^*(s_1)>V^*(s_2)$ entonces el agente preferirá el estado “s_1” sobre el estado “s_2”, ya que el primero de los estados tendrá una mayor recompensa futura acumulada. En los problemas libres de modelos la función de recompensa y la de transición son funciones desconocidas, no pudiéndose por tanto calcular de forma directa el valor óptimo $V^*(=V^{\pi^*})$.

Para abarcar este problema en un MDP, en donde las acciones y los estados son discretos pero la función de transición así como la función de recompensa son desconocidas se utiliza la función Q(s,a), que se define de la siguiente forma:

$$Q(s,a) \equiv \left[r(s,a)+\gamma V^*(\delta(s,a))\right] \tag{4}$$

Es decir, la suma de la recompensa de aplicar la acción "a" en el estado “s” junto con el valor (descontado por γ) obtenido al aplicar la política óptima al estado resultante de aplicar la acción “a” en el estado “s”.

Esta ecuación nos permite reescribir a π* de la siguiente manera:

$$\pi^*(s) \equiv \underset{a}{argmax}\, Q(s,a) \tag{5}$$

Esta forma de la ecuación muestra que, en caso de conocer la función Q en vez de la función V*, se puede también calcular la política óptima. Lo que se traduciría en que el agente puede, sin conocer a priori la función r(s,a) ni la función δ(s,a), seleccionar las acciones óptimas. Sólo se

consideraría entre las acciones posibles, estando el agente en el estado “s”, aquella que maximice el valor de Q (Mitchell, 1997).

Siendo que:

$$V^*(s) \equiv \max_{a'} Q(s, a') \tag{6}$$

Entonces se puede reescribir a Q(s,a) como:

$$Q(s, a) = [r(s, a) + \gamma \max_{a'} Q(\delta(s, a), a')] \tag{7}$$

Quedando de esta forma expresado el valor de Q en función de r y Q.

Si se denomina Q’ a un estimativo de la función Q. Debido a que los estados y las acciones son discretas, se puede pensar en cada acción y en cada estado como entradas independientes, las cuales definen una tabla en donde se pueden almacenar los valores estimados Q’ (uno por cada par (s,a)). A su vez, se puede determinar que cada valor de Q’ de la tabla sea inicializado en 0 (pudiendo ser tomados valores aleatorios por cada entrada, de forma equivalente).

Entonces, al ejecutar la acción "a" en el estado “s” el agente puede observar tanto la recompensa r=r(s,a) como el nuevo estado s’=δ(s,a). La actualización de los valores de la tabla que almacena el agente se pueden actualizar de la siguiente forma:

$$Q(s, a) \leftarrow [r + \gamma \max_{a'} Q(s', a')] \tag{8}$$

Logrando de esta forma actualizar el estimativo de Q’(s,a) aproximando este último de forma cada vez más precisa al verdadero valor de Q(s,a) (se puede probar que Q’ a la larga converge al verdadero valor de Q(s,a) tomando ciertas suposiciones, como, por ejemplo que cada dupla estado-acción sea visitada infinitamente, ver Mitchell (1997)).

Este tipo de aprendizaje (q-learning) tiene el inconveniente que aplica solamente a estados y acciones discretos, además de ser un proceso de aprendizaje relativamente lento. Es por esto que se ha sustituido en ciertos casos la tabla con entradas independientes por un red neuronal, en

donde lo que se actualiza son los pesos sinápticos de las “neuronas” de la red en vez de los valores del estimativo Q’ directamente.

2.7.2. Red neuronal artificial de base radial

Se han desarrollado distintas variedades de redes neuronales basadas en Q-learning, como por ejemplo, la red Q-Kohon (Santos and Touzet, 1999a) y la Red artificial que aplica una función de base radial implementando q-learning presentada por Santos y Touzet (1999b).

Esta última (Figura 6) está formada por tres capas: una capa de entrada, una capa oculta y una de salida. En la capa de entrada (vector entrada), se codifica el estado en que se encuentra el pez así como el valor estimado Q (el valor del estado suele, al menos en parte estar determinado por los sensores del agente) para el par (s,a). La capa de salida contiene las neuronas que integran la señal de salida, que pueden, por ejemplo, en el caso de un agente que tenga que navegar en un determinado espacio, definir los componentes motores de dicho agente. La capa oculta de esta red “contiene” en cierta medida la función de base radial con la cual se comparan los componentes del vector entrada con los pesos sinápticos de cada neurona de la capa oculta y posteriormente, en base a esto, se elige la neurona de la capa salida a disparar. La arquitectura de esta red está formada por neuronas completamente conectadas entre una capa y la siguiente, es decir, todos los elementos (neuronas) de la capa de entrada están conectados con cada uno de los elementos de la capa oculta y estos con los de la capa de salida. Otra característica importante de esta red es que es creciente, es decir, en determinadas circunstancias, el algoritmo se encarga de añadir una nueva neurona oculta a la red.

El tipo de funcionamiento de esta red es del tipo de winner-takes-all, ya que la red selecciona a la neurona de la capa oculta cuyos pesos de entrada son de mayor similitud con el vector entrada (aplicando una función de base radial) para seleccionar la acción a tomar, quedando dicha acción determinada por los pesos de salida de la capa oculta. Esta red, mediante una función de recompensa, asigna penalizaciones o recompensas a aquellas acciones que, respectivamente, determinen estados no deseables o deseables. El “objetivo” de esta red es seleccionar aquellas

acciones que permitan generar un mayor valor de recompensas acumulado (tomando en consideración el descuento para las acciones más alejadas temporalmente), pero en lugar de seleccionar directamente entre estados se selecciona entre neuronas de la capa oculta que representan, a un conjunto de estados con sus valores Q asociados. La función que se utiliza para evaluar el parentesco entre el vector de entrada y los pesos de entrada de la capa oculta es una función similar a una gaussiana. Esta red "selecciona" a aquella neurona cuyo parentesco entre los pesos de entrada y el vector entrada (que codifica un estado) sea mayor. Además, dicha red contiene un valor umbral cuya función es determinar si el grado de similitud entre el vector de entrada y los pesos de la neurona oculta de mayor parentesco es aceptable[2]. En caso de serlo la acción a ser tomada por el agente será equivalente a los pesos de salida de tal neurona multiplicado por un factor que se puede interpretar como inherente a la capa de salida. En caso contrario se agrega una neurona oculta conectada con todas las neuronas de entrada y de salida, cuyos pesos de entrada serán iguales al vector entrada y los pesos de salida serán elegidos de forma aleatoria (Figura 7). En la Figura 8 se muestra el algoritmo de actualización del valor de Q-max, que se utilizará para actualizar los pesos relacionados con el valor Q (Figura 9). En la Figura 9 se describe el algoritmo de actualización de los pesos sinápticos de entrada (w_s y w_Q) y de salida (w_a).

[2] La aceptabilidad es en cierta medida arbitraria, pudiendo variar de acuerdo a las necesidades del operador. Esta es dependiente del umbral de aceptación.

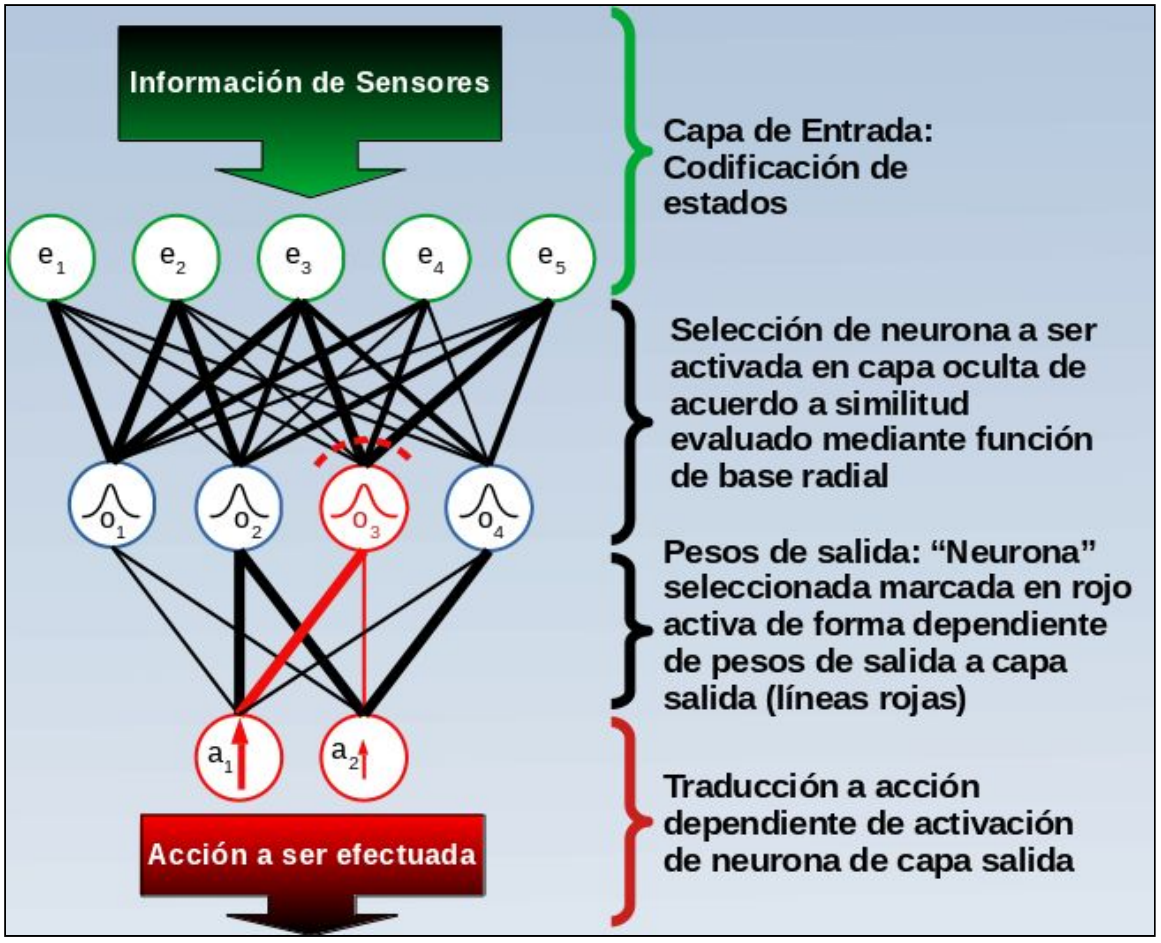

Figura 6. Ejemplo de arquitectura de red neuronal basada en una función radial. e_i son las "neuronas" correspondientes a la capa de entrada. o_i son las "neuronas" de la capa oculta las cuales contienen cada una un prototipo con el cual comparar el vector de entrada (dado por los pesos de entrada representados en la imagen por el grosor de las líneas), utilizando como comparador una función de base radial. La salida de esta red será una combinación lineal de cada neurona efectora (a_i) que va a tener una activación dependiente de los "pesos de salida" de la neurona de la capa oculta seleccionada.

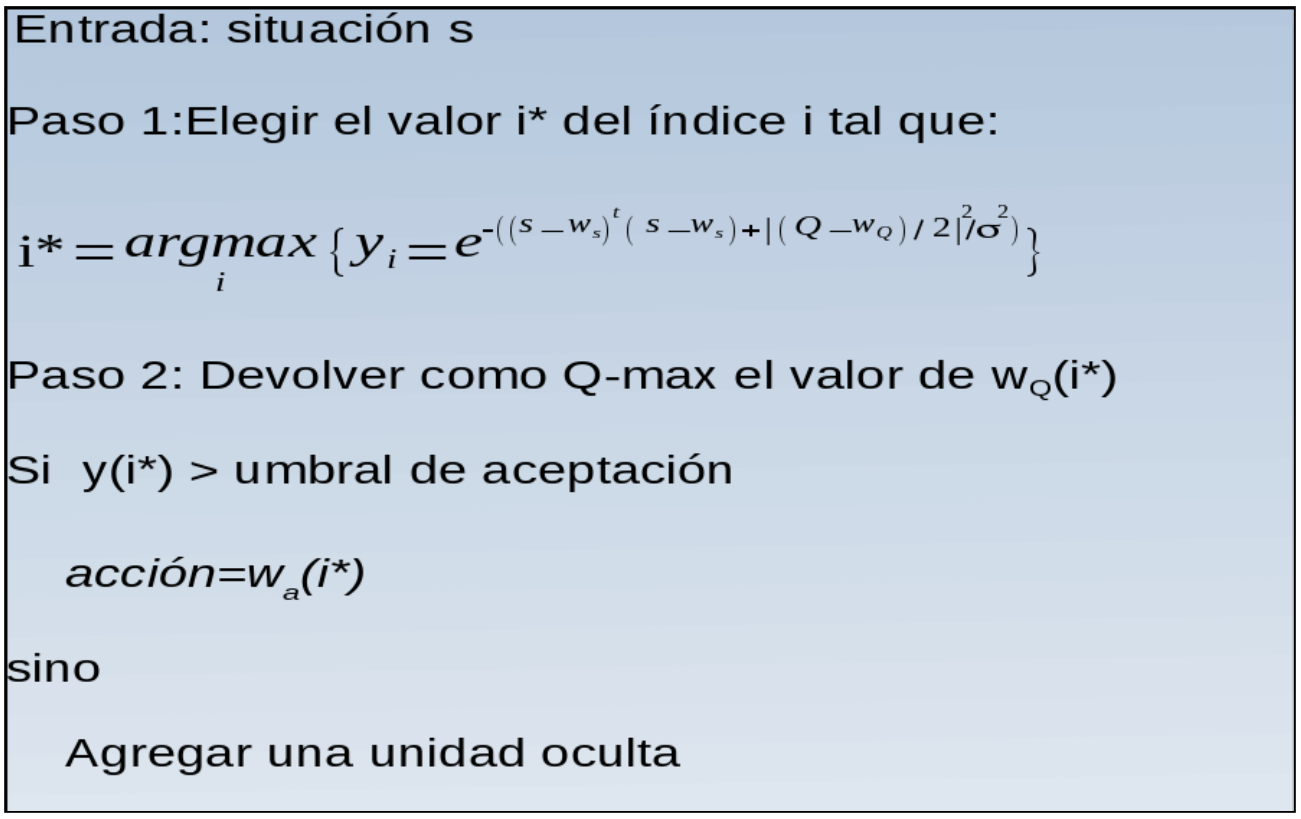

Figura 7. Algoritmo de actualización de parámetros en un algoritmo RBFNN clásico. w_s, w_Q y w_a corresponden a los pesos entrada que conecta las neuronas entrada con una neurona oculta (i), los pesos entrada que conecta el valor Q de entrada con una neurona oculta (i) y los pesos salida de la neurona oculta(i) de manera respectiva. i corresponde a la i-ésima neurona de la capa oculta. Modificado de Santos y Touzet (1999b).

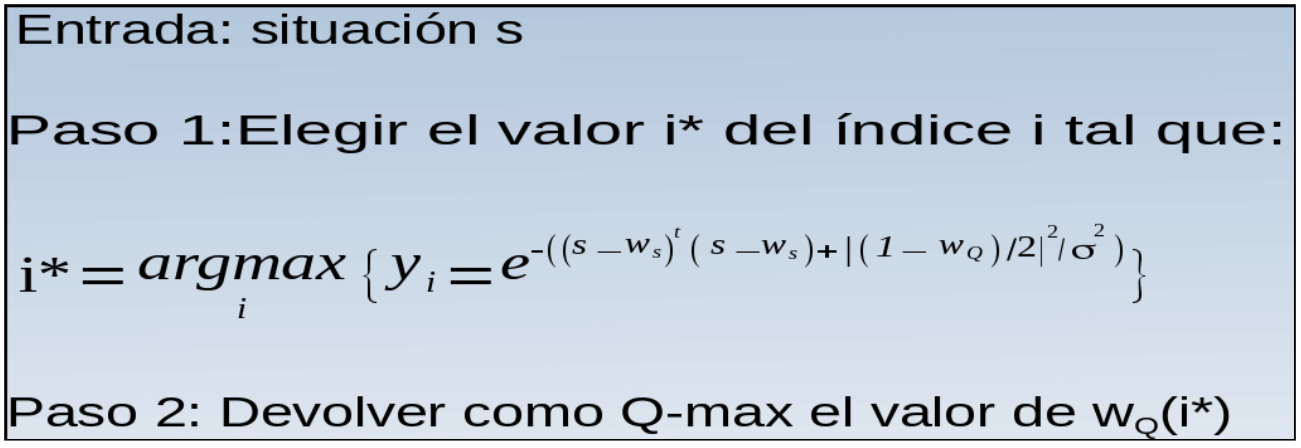

Figura 8. Algoritmo de selección del valor máximo de Q (Q-max) mediante la utilización de una función de base radial. El valor 1 (de la función y_i) puede ser otro y corresponde al valor de

entrada Q que sea elegido, en general se trata de maximizar dicho valor de forma que el valor Q-max sea lo más elevado posible. w_Q es el peso de entrada de la neurona Q a la neurona oculta (i-ésima). El peso $w_Q(i^*)$, será el valor elegido como Q-max por ser, en general el comparativamente más cercano al máximo y al estado según la función y_i. Modificado de Santos y Touzet (1999b).

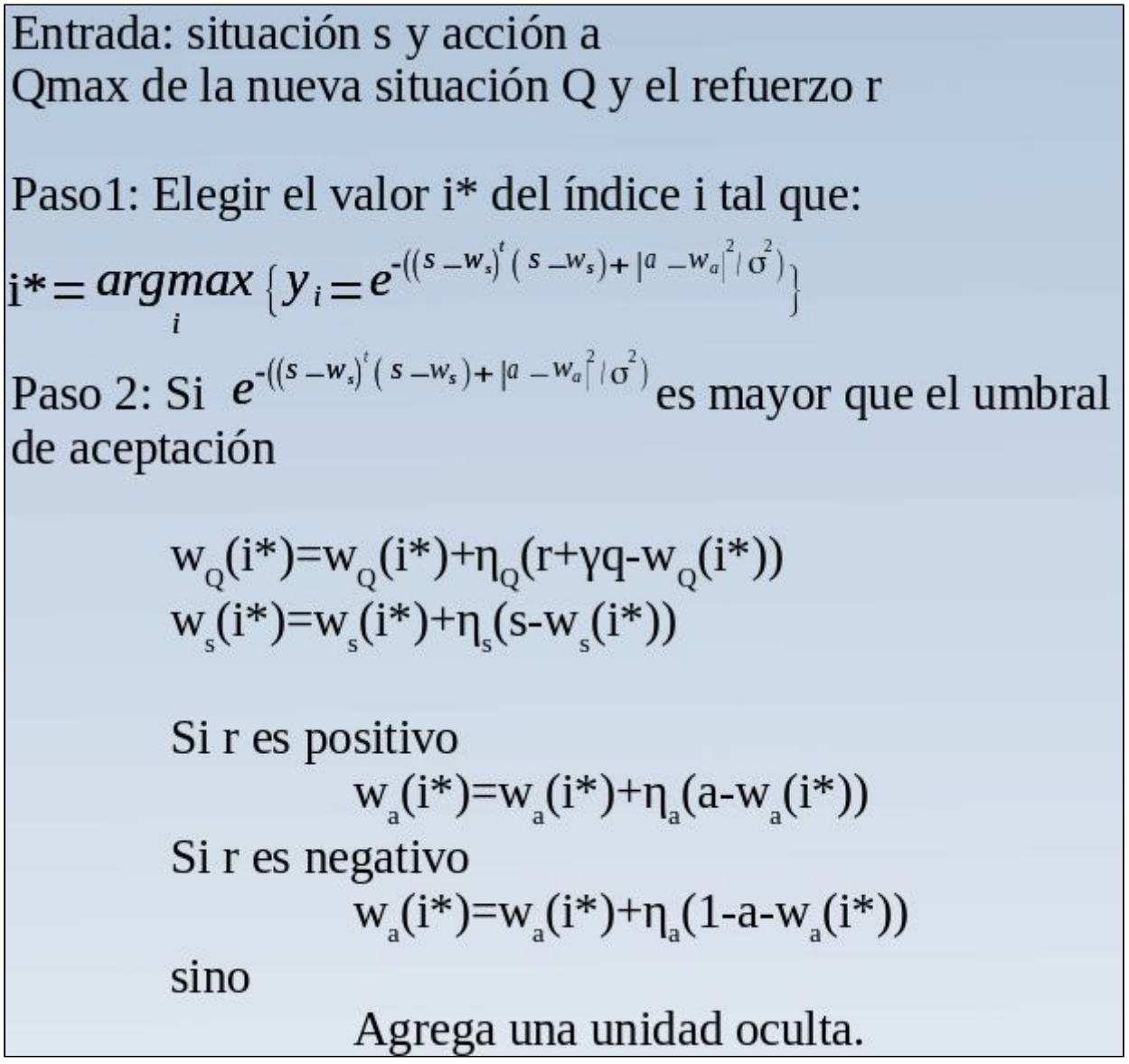

Figura 9. Actualización de los pesos sinápticos de la red neuronal. w_Q y w_S corresponden a los pesos de entrada de las neuronas ocultas provenientes del elemento de entrada con el valor Q y del vector "s" (estado). En el que cada elemento del vector "s" corresponde al valor de cada neurona de la capa de entrada. η_Q, η_s y η_a son constantes que toman valores entre 0 y 1

(constantes de la velocidad de aprendizaje de la red). Este algoritmo actualiza no solamente los pesos sinápticos relacionados con Q (w_Q) sino que también actualiza los pesos sinápticos relacionados con el estado y la acción (w_s y w_a). Modificado de Santos y Touzet (1999b).

3. Objetivos generales y objetivos específicos

3.1. Objetivos generales

En el presente estudio se tratará de analizar el comportamiento de acercamiento al objeto analizando de forma heurística este comportamiento. Tratando de responder a las siguientes preguntas:

- Un agente que efectúa un aprendizaje basado en incrementar la imagen eléctrica ¿logra efectuar de manera emergente una corrección dependiente del ángulo, que se forma entre el vector distancia y el vector que parte del centro hacia la cabeza del pez, al recompensar el incremento del pico de la imagen?
- En caso de que exista: ¿Tal corrección efectuada por el agente es suficiente como para lograr un alineamiento dependiente de la distancia?

Analizando para esto si (a partir de recompensar o penalizar acciones que favorezcan o dificulten, respectivamente, el incremento del pico de la imagen eléctrica) existe una correlación entre la amplitud, así como el sentido, del ángulo de giro efectuado por el agente y el alineamiento entre el pez y el objeto. Así como si existe una correlación entre dicho alineamiento y la distancia.

Lo que, en caso afirmativo, permitirá responder si:

- ¿Este sistema dotado de un mecanismo de aprendizaje por refuerzo es capaz, basado simplemente en recompensar el incremento del pico de la imagen eléctrica, de converger sobre un modo de funcionamiento similar al modelo sencillo y “mínimo” denominado “Toymodel”?

3.2. Objetivos específicos

1. Construir un modelo simplificado del pez que navegue en un espacio de 2 dimensiones y que posea un sistema sensorial a partir del cual obtiene una imagen del entorno

(compuesto básicamente por un objeto que se supone conductor ubicado en una posición aleatoria dentro de ese espacio).

2. Implementar el algoritmo de aprendizaje por refuerzo ya mencionado, recompensando el aumento de la amplitud de la imagen eléctrica y penalizando la disminución.
3. Realizar simulaciones para verificar si se obtiene un patrón de comportamiento cualitativamente consistente con lo observado experimentalmente.
4. Estudiar las variaciones del rendimiento frente a variaciones de los parámetros relevantes con el fin de obtener un panorama acerca del rendimiento del programa.
5. Analizar si el modelo aplicado en este estudio presenta algún patrón de corrección angular relativo al ángulo similar al realizado por el “Toymodel” y, en caso de que este exista, si es suficiente para obtener el alineamiento del pez respecto de la distancia.

4. Materiales y Métodos

4.1. Configuración del agente y su arena

Los experimentos se realizaron utilizando el programa “GNU Octave” (John W. Eaton & Wehbring, 2015) utilizando el algoritmo Q-RBFNN (utilizado por Santos y Touzet (1999b) del inglés Radial Basis Function Neural Network). En los experimentos realizados se definió el espacio de forma bidimensional, esto es, se consideró que el pez se encontraba en aguas muy llanas y que el pez únicamente podía tomar acciones en el sentido de avance. Más precisamente, este vector de avance se generó por la suma vectorial de dos vectores de módulo variable con igual sentido en la componente horizontal y sentidos opuestos en la componente del eje vertical (si se toma como eje horizontal al eje formado entre la cola y la cabeza del “pez”), el ángulo formado entre el eje horizontal y cada vector se determinó como π/3 y -π/3 (ver Figura 10). El número de sensores

utilizado fue de 7: uno en la posición de la cabeza (la posición de la cabeza en realidad se determinó como este sensor) y 3 pares de sensores laterales (cada par dispuesto de forma simétrica con respecto al eje formado entre la cabeza y la cola). El campo eléctrico sensado por el pez (luego de aplicar una descarga) fue simplificado de forma tal que el pez en vez de recibir una señal que cae con la cuarta potencia de la distancia cae de forma lineal con la distancia. El objeto se simplificó a un punto, para reducir la complejidad del problema. La red que se simuló, utilizada por el pez, está compuesta de 8 neuronas de entrada, 2 neuronas de salida e inicialmente de 2 neuronas ocultas. Los pesos de entrada de cada neurona oculta se inicializan en 0 y los valores de los pesos de salida se determinaron de forma aleatoria. El vector entrada, los pesos de las neuronas de la capa oculta (tanto de entrada como de salida) y el vector salida presentan valores comprendidos entre 0 y 1. Esto no se cumple para el valor q de entrada y los valores w_Q de cada neurona oculta que, a priori, pueden ser un poco superiores. Quedando los valores de los elementos del vector entrada (a excepción del octavo que corresponde al valor q) codificados como el inverso de la distancia entre el sensor de cada neurona de entrada y el objeto, sobre el valor máximo registrado entre todos los sensores de esa misma distancia, es decir:

$x_i=d_i^{-1}/\max(d_i^{-1})$ i=1, 2,…, 7

En donde d_i es el valor de la distancia de cada sensor "i" al objeto y x_i es cada uno de los valores de los sensores.

El algoritmo aplicado es, como ya se mencionó, en general igual al utilizado por Santos y Touzet (1999b), los parámetros iniciales fueron variados de forma de poder comparar los resultados obtenidos (evaluando por ejemplo el número de iteraciones en la que el agente convergía al resultado deseado en relación a la distancia, el número de recompensas positivas en relación al número total por episodio[3], etc.). La función de recompensa (FR) quedó determinada por tanto, como:

[3] Una episodio está formada por todas aquellas iteraciones desde que al pez se le asigna una posición aleatoria hasta que llega al objeto o hasta que llega a la iteración máxima (anexo a1).

$$FR = \begin{cases} +1 & si \quad max(x_i) - max(x_{i-1}) > 0 \\ -1 & si \quad max(x_i) - max(x_{i-1}) < 0 \\ 0 & en \quad otro \quad caso \end{cases} \tag{9}$$

Esto se traduce como: si el pez obtiene una imagen eléctrica con un pico mayor al pico de la imagen anterior, entonces el agente es premiado de forma positiva, en caso que sea menor se penaliza al agente y en otro caso (es decir que el pico de la imagen sea de similar amplitud) se asigna un refuerzo nulo. Esta forma de definir a la función de recompensa fue elegida, ya que el aumento de la imagen eléctrica supondría un acercamiento del pez al objeto, si la imagen eléctrica se ve reducida supondría un alejamiento y en otro caso podría suponerse que la distancia se mantiene constante. Implícitamente se está suponiendo que el pez realiza la misma cantidad de descargas eléctricas como iteraciones tiene el programa (ya que el agente realiza una medida por iteración y en base a eso elige su consecuente acción[4]).

El algoritmo ya descrito se definió para que, de forma iterativa, actualizara los valores de la red y en caso que la recompensa recibida sea nula, o que, a la hora de generar una acción no se supere el umbral de aceptación, se agregue una nueva neurona en la capa oculta (Figura 7). Generando en el último caso una acción de carácter meramente aleatorio y que después es ajustada de forma iterativa de acuerdo al algoritmo ya definido.

El valor mínimo y máximo de cada uno de los vectores de salida fueron (los vectores que terminan determinando la acción), respectivamente, de 0 y 1 en todas las simulaciones generadas.

Los parámetros se definieron para todos los experimentos (a excepción de cuando éstos eran los que estaban siendo estudiados) en 0,3 el descuento gamma (γ); 0,5 el valor de la constante de la velocidad de aprendizaje de los pesos sinápticos w_Q (η_Q); 0,01 la constante de actualización de los pesos de los estados w_S (η_s); 0,4 la constante de la velocidad de aprendizaje para los pesos de salida w_a (η_a); 0,4 a sigma de la función de base radial (σ); 0,75 el valor umbral y 1,3 el valor-Q.

La arena en la cual se posicionó al agente se definió como infinita, es decir el pez no tenía paredes que limitaran su movimiento. Las posiciones iniciales tanto del pez como del objeto se

[4] Una iteración podría pensarse como un época, ya que por iteración ocurre un ciclo de aprendizaje. Esto es, toma una acción, recibe un refuerzo además de obtener un nuevo estado y a partir de esto actualiza los valores de la red.

determinaron de forma aleatoria dentro de un rango no mayor a 2 unidades de distancia. El diámetro del pez (la longitud) fue de 0,16 unidades de distancia.

4.2. Parámetros

Se realizó un estudio parcial del espacio de rendimiento del programa de acuerdo con los distintos parámetros. Para lograr ello se analizaron en conjunto valores de Q de entrada (en el algoritmo original se dejaba en 1) y el valor del umbral de aceptación (en el caso del valor Q, se tomaron valores entre 0,6 y 1,3 variando en pasos de 0,1 unidades, en el caso del umbral se evaluó desde 0,45 hasta 0,85). También se analizó de forma similar los espacios formados por la variación de los parámetros de η_a (de 0,05 a 0,95 en pasos de 0,1 unidades) y η_s (de 0,01 a 0,51 variando en pasos de 0,05 unidades); η_Q y γ (ambos variando en 0,1 unidades entre 0,05 y 0,95). De esta forma se pudo obtener un perfil de rendimiento del espacio formado por cada pareja de parámetros). Para estudiar esto, se emplearon para cada simulación 500 episodios (de aprendizaje) y se analizaron (se calculó el promedio y el error estándar respectivo) en conjunto y también por separado los primeros 50 episodios y los últimos 450, esto para determinar si el comportamiento medio del pez mejoraba o empeoraba, en términos generales para los parámetros variados[5]. Para cada caso se obtuvieron tanto los cocientes: Nº total de recompensas sobre Nº iteraciones (R+/it) como el N°iteraciones sobre distancia (it/dist), con lo cual se pudo estudiar de forma cualitativa (al graficar ambos cocientes) la relación entre ambos. El rendimiento del programa se evaluó en base al primer cociente, mientras que el segundo sirve como un estimativo de cuán “rápido” llega al objeto (si es que llega, hay que recordar que ambos cocientes están limitados en el número de iteraciones, ya que estas no pueden superar el valor de iteración máximo (anexo a1)). El hecho de comparar ambos cocientes ofrece la posibilidad de estimar si hay cierta optimización en el número de pasos efectuados por el agente a medida que aumenta el rendimiento R+/it.

---

[5] Por cada valor de cada parámetro variado se efectuó una simulación constituida por 500 episodios de aprendizaje.

4.3. Evaluación rendimiento a corto y mediano plazo

También se estudió la distribución del ya descrito rendimiento promedio, fijando los parámetros (en los valores que ya se definieron como estándar, basándose en su mayoría en Santos y Touzet, 1999) y realizando unas 100 simulaciones cada una conformada por 100 episodios de aprendizaje, de forma de poder observar el comportamiento del programa en un corto y mediano plazo. Junto con esto se graficaron algunas de las trayectorias realizadas por el agente (en distintos episodios de aprendizaje), con el fin de poder ilustrar el patrón de movimiento del mismo (Figura 11).

4.4. Alineamiento medio en relación a la distancia

Para estudiar el alineamiento del pez con respecto al objeto, se midió el ángulo (θ) formado entre el vector que va desde la cabeza a la cola del pez junto con el vector que va desde el centro del pez hacia el centro del objeto, esta vez se realizaron unas 100 simulaciones con 500 episodios de aprendizaje. De estos 500 se evaluaron los últimos 250 episodios de aprendizaje de forma de minimizar, dentro de lo posible, los comportamientos aleatorios del pez, producto del aprendizaje de nuevas políticas. Estos ángulos se analizaron con respecto a las diferentes distancias. Para lograr esto último se registraron las iteraciones que pertenecían a un determinado intervalo de distancia y posteriormente se promediaron[6] los ángulos que correspondían a dicho intervalo (la longitud de los intervalos se obtuvo de la diferencia entre la distancia máxima y la mínima sobre el número de intervalos, siendo este de 100). Además se hizo el mismo análisis pero solamente con aquellos casos que lograron alcanzar el objetivo (casos exitosos), es decir de aquellos 250 episodios de aprendizaje se seleccionaron solamente aquellos que efectivamente lograron alcanzar el objetivo. Con respecto al ángulo θ se midió escalándolo de π a -π (de acuerdo a si el

[6] A la hora de realizar los promedios arriba descritos se tomó en realidad el valor absoluto de dichos ángulos, ya que un promedio simple no sería indicativo de alineamiento (esperándose un valor promedio de 0 para trayectorias aleatorias).

objeto estaba a la derecha o a la izquierda del pez, respectivamente. El valor 0 corresponde a la cabeza y a medida que se aleja de la cabeza hacia zonas más posteriores el valor en términos absolutos se incrementa hasta π, valor correspondiente a la cola del pez).

4.5. Ángulo de giro (Φ) en relación al ángulo de alineación (θ)

Además de esto se determinó el ángulo de giro de una iteración a la siguiente evaluando este ángulo de acuerdo con el ángulo θ. El ángulo de giro (Φ) se definió como:

$$\Phi(i)=\begin{cases} \Delta\theta(i) & si \quad |\Delta\theta(i)| < \Pi \\ 2\Pi - \Delta\theta(i) & si \quad \Delta\theta(i) > \Pi \\ 2\Pi + \Delta\theta(i) & si \quad \Delta\theta(i) < -\Pi \end{cases} \tag{10}$$

siendo $\Delta\theta(i) = \theta(i+1) - \theta(i)$ (11) e i una iteración.

De esta forma, el signo de Φ(i) indica si el sentido de rotación fue horario o antihorario (es decir, positivo y negativo, respectivamente).

Para evaluarlo de acuerdo al ángulo θ se promediaron los distintos valores de Φ(i) para los cuales el ángulo θ(i) pertenecía a un determinado intervalo (estando la longitud de cada intervalo determinada por la distancia entre el mayor y el menor de los ángulos (π y -π, aproximada y respectivamente) sobre el total de intervalos, que fue de 100). Además se obtuvo para cada uno de los intervalos el número de datos (número de distintas épocas) que cada uno contenía.

Estos datos se utilizaron para averiguar si se lograba obtener una corrección angular para cada amplitud angular acorde con la implementada por Gómez-Sena, es decir una función de corrección angular dependiente de θ que sea de tipo sigmoide.

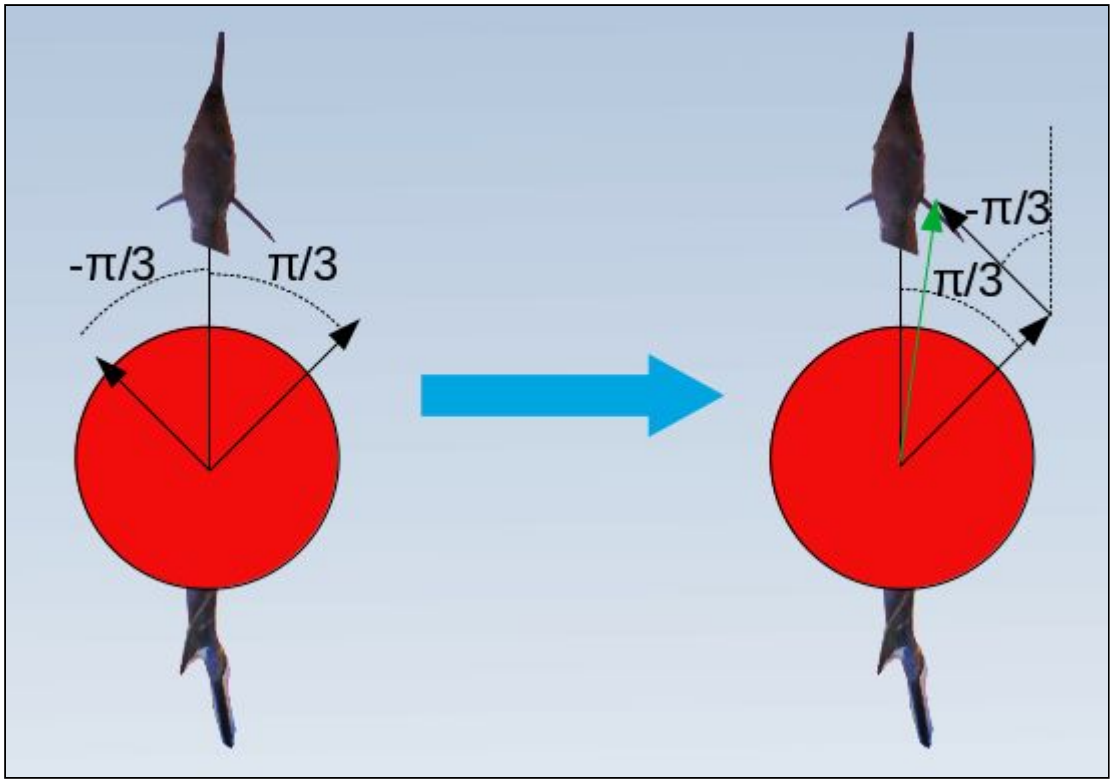

Figura 10. Representación esquemática de los vectores (flechas negras) que definen al vector desplazamiento (flecha verde) obtenido mediante la suma de estos dos vectores (ilustración de la derecha). Ambos vectores son de módulo variable pero con ángulos fijos en relación a la cabeza (-π/3 y π/3, para vector con componente izquierda y derecha, respectivamente).

5. Resultados

5.1. Observaciones cualitativas del comportamiento del agente

En la Figura 11 se muestran ejemplos de las distintas trayectorias efectuadas por el pez, en las que sí logró acercarse al objeto. Al observar el comportamiento de forma cualitativa se observó que el agente en general se acercaba a este y que al estar a cierta distancia muy cercana al mismo el agente solía oscilar de forma mayor a lo que se hubiese esperado en un principio.

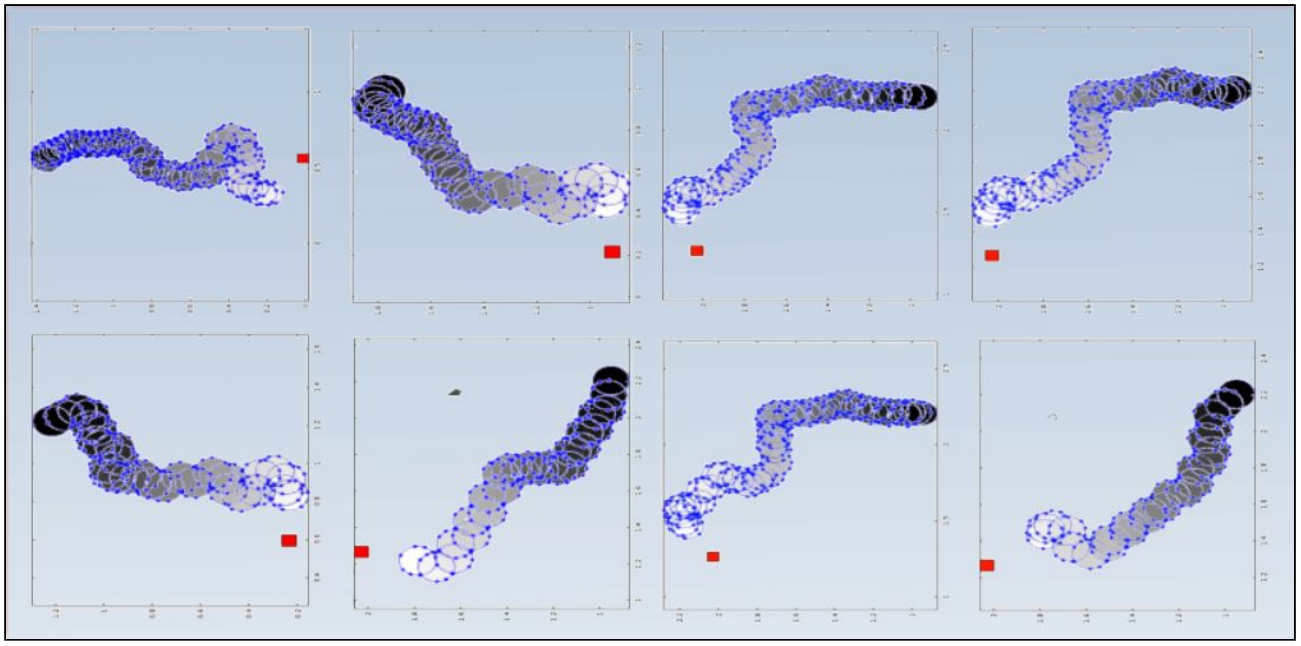

Figura 11. Ejemplos de trayectorias exitosas: El agente está representado por un círculo, sus sensores por asteriscos azules. En escala de grises se muestra el orden temporal de cada iteración (las tonalidades más claras corresponden a eventos más recientes temporalmente). En rojo se dibujó al objeto como un cuadrado (aunque en el modelo se definió como un punto).

5.2 Rendimiento a corto y mediano plazo

El patrón general registrado para el caso en el que se midió el rendimiento por iteración con los valores de parámetros definidos como estándar (ver métodos), mostró un incremento en los valores medios, en general, a medida que el pez efectúa cada vez más episodios (ver Figura 12). Esto ocurrió hasta aproximadamente el episodio número 25, en donde a partir de ahí se estabilizó el número de recompensas promedio por iteración en un entorno que varía aproximadamente entre 0,725-0,825. Siendo además el primer valor del rendimiento promedio registrado para el primer episodio cercano a 0,65 R+/it.

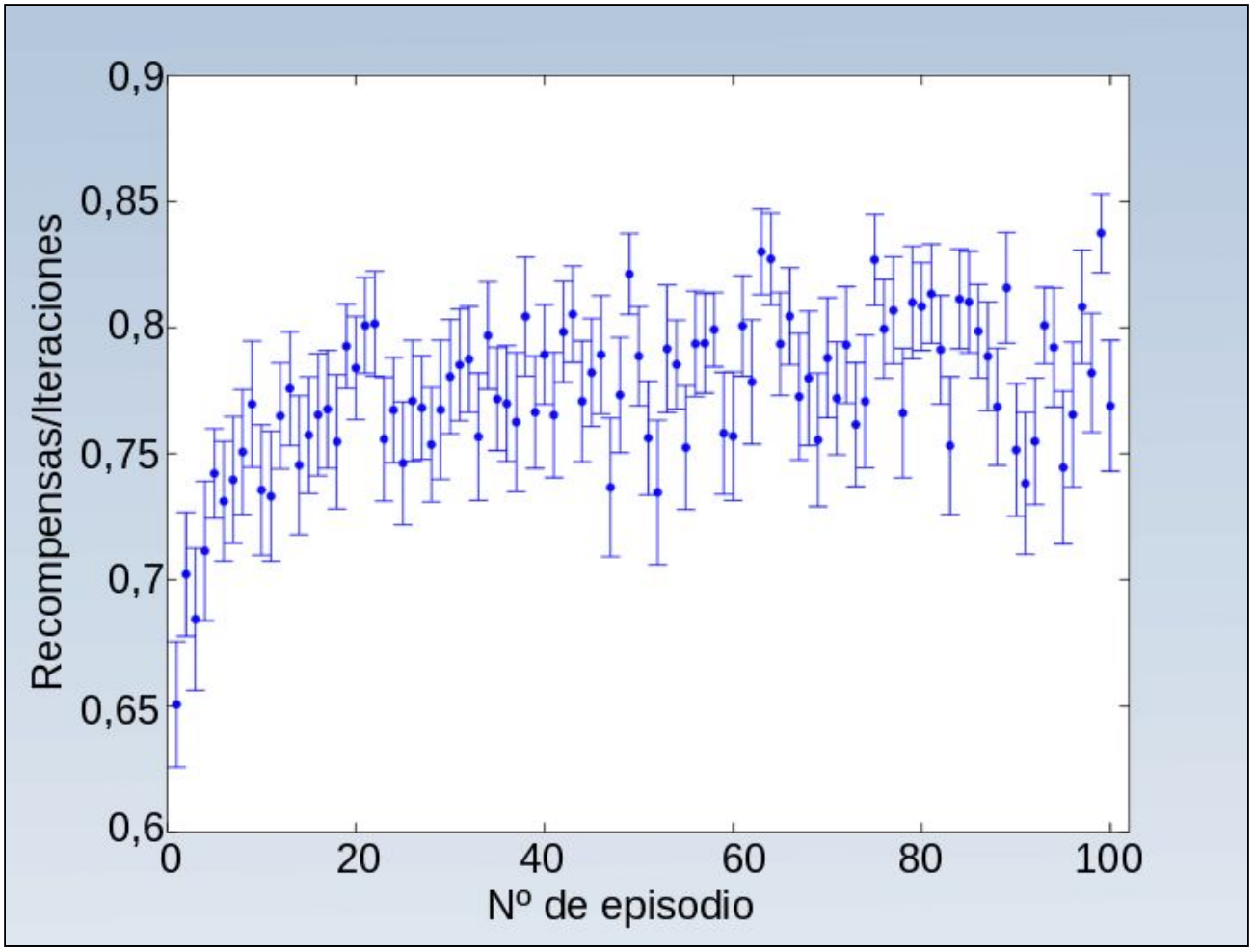

Figura 12. Promedio de recompensas por iteración registrado para cada episodio de aprendizaje promediando para 100 casos. Las barras corresponden al error estándar calculado.

5.3. Parámetros

Lo que se observó para el caso de la variación del valor-Q conjunto con el valor umbral fue que el valor mínimo de recompensas positivas por iteración fue de 0,2160 (+/-0,001) y se registró en el valor-Q de 0,7 y el valor del umbral de 0,45. El valor máximo fue de 0,8652(+/-0,0005) R+/it y se registró en el valor del umbral de 0,65 y en el valor-Q de 1,3. Se observó además un patrón de aumento en el rendimiento medio a medida que se incrementa, en términos generales, el valor-Q y el umbral de aceptación (Figura 13, izquierda: arriba, medio, abajo).

En general, se registró una política con un valor superior a 0,75 R+/it para aquellos parámetros iguales o mayores a 1,2 para el valor-Q y mayores o iguales a 0,65 para el valor umbral. Se pudo visualizar en este caso, además, que a medida que hubo un mejoramiento en el rendimiento (R+/it), este se pudo asociar con un buen desempeño del programa (un número elevado de recompensas positivas por iteración), mostrando que, más allá de la variación en el rendimiento presentada al variar este par de parámetros, cuando ambos eran elevados existió un aprendizaje que se puede considerar adecuado, en general.

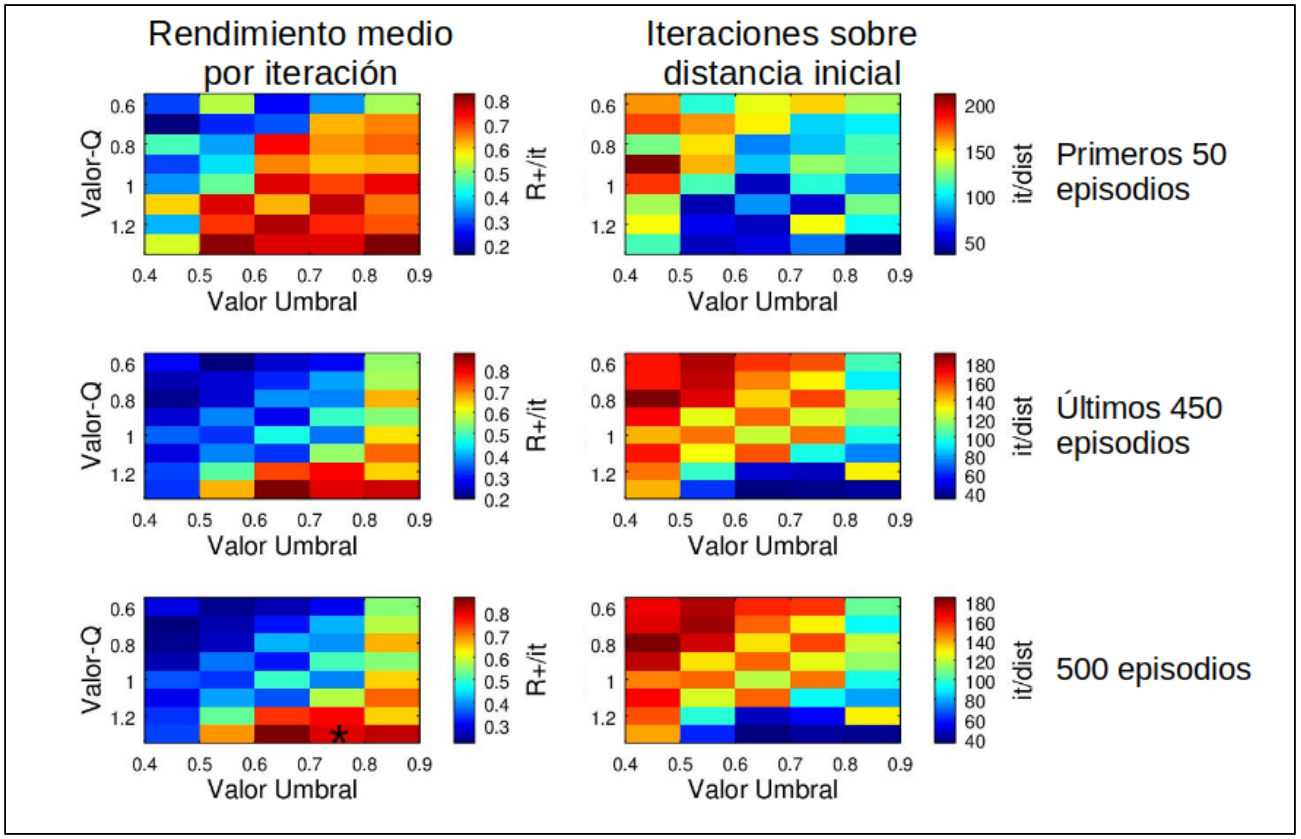

Figura 13. A la izquierda: Arriba: Rendimiento medio por iteración (R+/it) de los primeros 50 episodios en función del valor de Q y del valor umbral. Centro: Rendimiento medio por iteración (R+/it) de los últimos 450 episodios para cada valor de Q y el valor umbral calculado; Abajo: Rendimiento medio por iteración (R+/it) de los 500 episodios para cada valor de Q y del valor umbral calculado. Con un asterisco se señala la posición aproximada de los parámetros que se definieron como estándar. A la derecha: Arriba: Número de iteraciones sobre distancia (it/dist) para los primeros 50 episodios al variar los valores de Q y del umbral; Centro: Número de iteraciones sobre distancia (it/dist) para los últimos 450 episodios al variar los valores de Q y del umbral; Número de iteraciones sobre distancia (it/dist) para los 500 episodios al variar los valores de Q y del valor umbral.

En el caso donde se variaron los parámetros η_s y η_a (constantes de la velocidad de aprendizaje para los valores de los pesos w_s y w_a, respectivamente), se pudo visualizar el mayor rendimiento en los valores de η_a de 0,85 y 0,11 de η_s, el cual corresponde al valor de 0,8501+/-0,0005 R+/it. El valor mínimo fue de 0,5588+/-0,0009 y se registró en los valores de η_a de 0,45 y de η_s de 0,11. No obstante, no se registró una incidencia real en el rendimiento a excepción de algunos casos en los que el rendimiento fue relativamente bajo para valores elevados de η_s y para valores elevados de η_a donde este, en general, no fue aceptable (anexo a2).

En resumen, no se registró un patrón claro de distribución de los rendimientos, teniendo una distribución casi que aleatoria, excepto para valores elevados de η_s y η_a donde los valores R+/it registrados fueron menores.

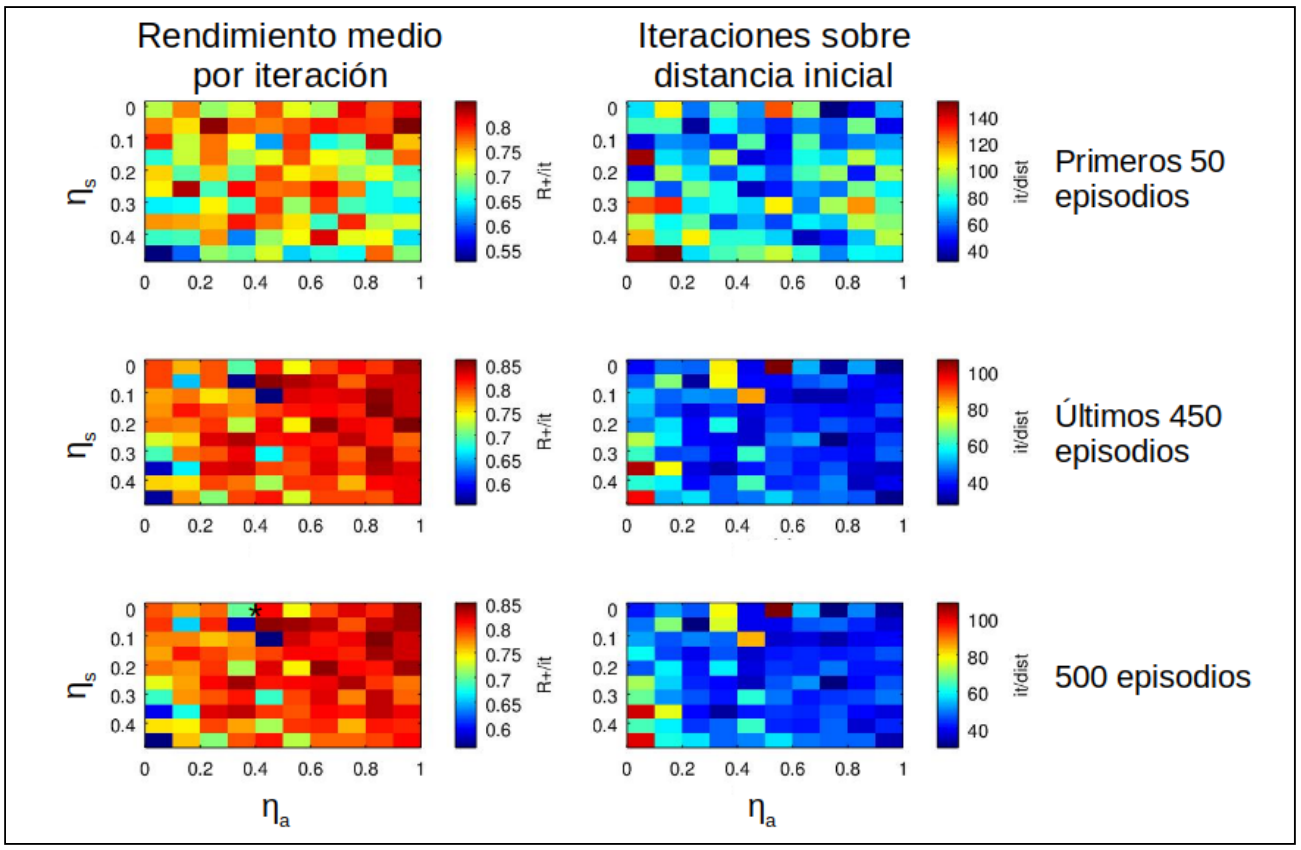

Figura 14. A la izquierda: Arriba: Rendimiento medio por iteración (R+/it) de los primeros 50 episodios para cada valor de η_s y de η_a calculado; Centro: Rendimiento medio por iteración (R+/it) de los últimos 450 episodios para cada valor de η_s y de η_a; Abajo: Rendimiento medio por iteración (R+/it) de los 500 episodios para cada valor de η_s y de η_a. Con un asterisco se señala la posición aproximada de los parámetros que se definieron como estándar. A la derecha: Arriba: Número de iteraciones sobre distancia (it/dist) para los primeros 50 episodios al variar los valores de η_s y de η_a; Centro: Número de iteraciones sobre distancia (it/dist) para los últimos 450 episodios al variar los valores de η_s y de ηa; Número de iteraciones sobre distancia (it/dist) para los 500 episodios al variar los valores de η_s y de η_a.

Al variar el factor de descuento (γ) y la constante de la velocidad de aprendizaje para los valores de los pesos w_Q (η_Q) (ambos implicados en la ecuación de actualización de los pesos w_Q), el valor máximo fue de 0,8516+/-0,0005 R+/it y se alcanzó en el valor de γ de 0,25 y el valor η_Q de 0,15. Mientras que el valor mínimo registrado fue de 0,1955+/-0,0006 R+/it y se alcanzó en el valor de γ de 0,05 y de η_Q de 0,45.

Registrándose un patrón de mejores políticas para aquellos valores bajos de γ y para valores elevados de η_Q (ver Figura 15). Encontrándose, exceptuando un caso (0.45 η_Q y 0.45 γ), todos los valores de rendimiento superiores a 0,75R+/it para valores de γ menores o iguales 0,35.

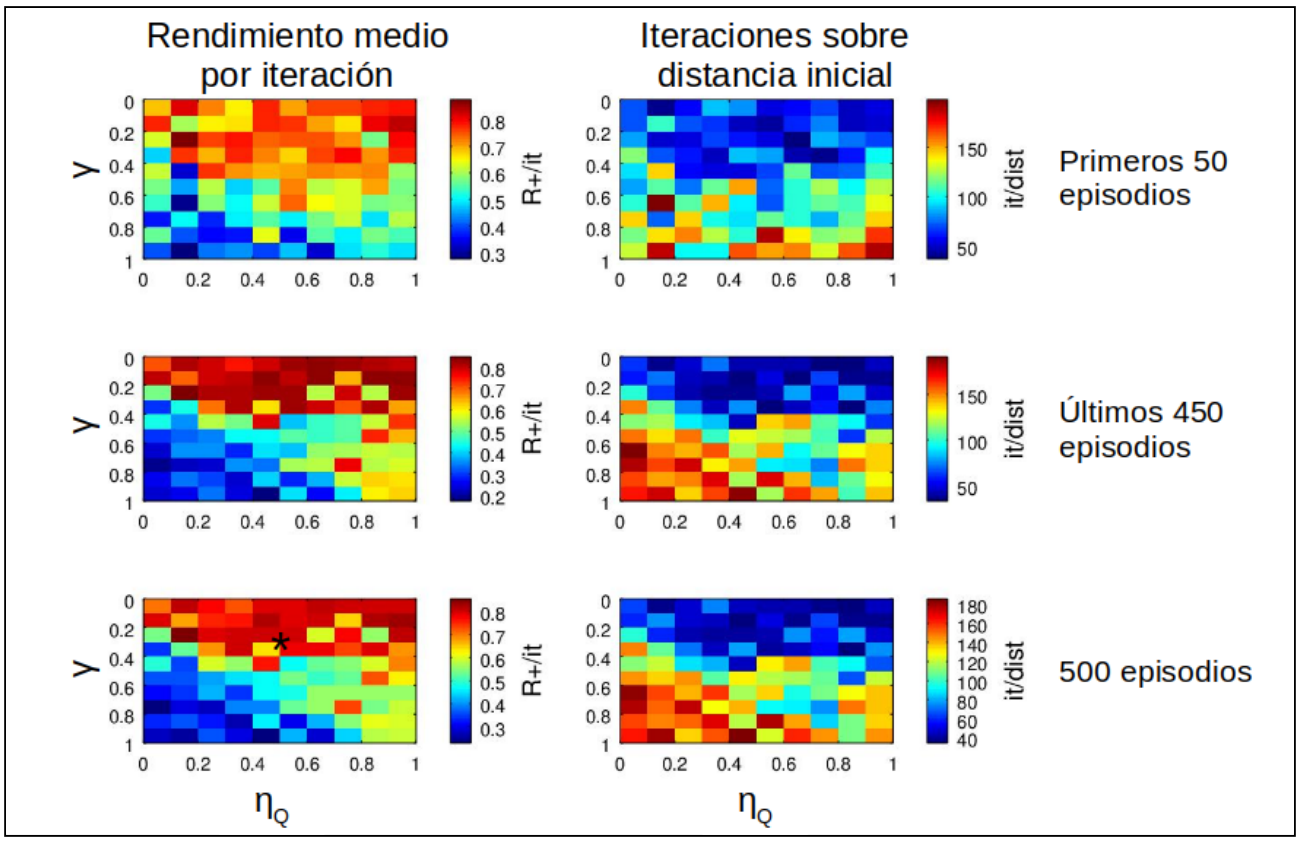

Figura 15. A la izquierda: Arriba: Rendimiento medio por iteración (R+/it) de los primeros 50 episodios para cada valor de η_Q y γ calculado; Centro: Rendimiento medio por iteración (R+/it) de las últimas 450 episodios para cada valor de η_Q y γ calculado; Abajo: Rendimiento medio por iteración (R+/it) de los 500 episodios para cada valor de η_Q y γ calculado. Con un asterisco se señala la posición aproximada de los parámetros que se definieron como estándar. A la derecha: Arriba: Número de iteraciones sobre distancia (it/dist) para los primeros 50 episodios al variar los valores de η_Q y γ; Centro: Número de iteraciones sobre distancia (it/dist) para los últimas 450 episodios al variar los valores de η_Q y γ; Número de iteraciones sobre distancia (it/dist) para los 500 episodios de entrenamiento al variar los valores de η_Q y γ.

Para el caso de σ y η_Q (respectivamente, constante relacionada con la ecuación de comparación, que está implicada en la capacidad de generalización de la red y constante de la velocidad de aprendizaje para los valores de los pesos w_Q), el valor de rendimiento máximo que se registró fue de 0,8640+/-0,0005 R+/it. Este se alcanzó en los valores de η_Q y σ de 0,65 y 0,55, respectivamente. El mínimo valor que se registró fue de 0,1605+/-0,0007 y se localizó en los valores de η_Q de 0,05 y de σ de 0,65.

El patrón que se observó consiste de una región en la que, todos los valores cuyo rendimiento fueron mayores que 0.75 R+/it se registraron para valores de sigma comprendidos entre 0.25 y 0.55 (ver Figura 16). Además de ser menos dependiente de η_Q, donde la mayor incidencia de este último se observa para valores bajos del mismo, registrándose en el caso para valores más extremos políticas con peor rendimiento.

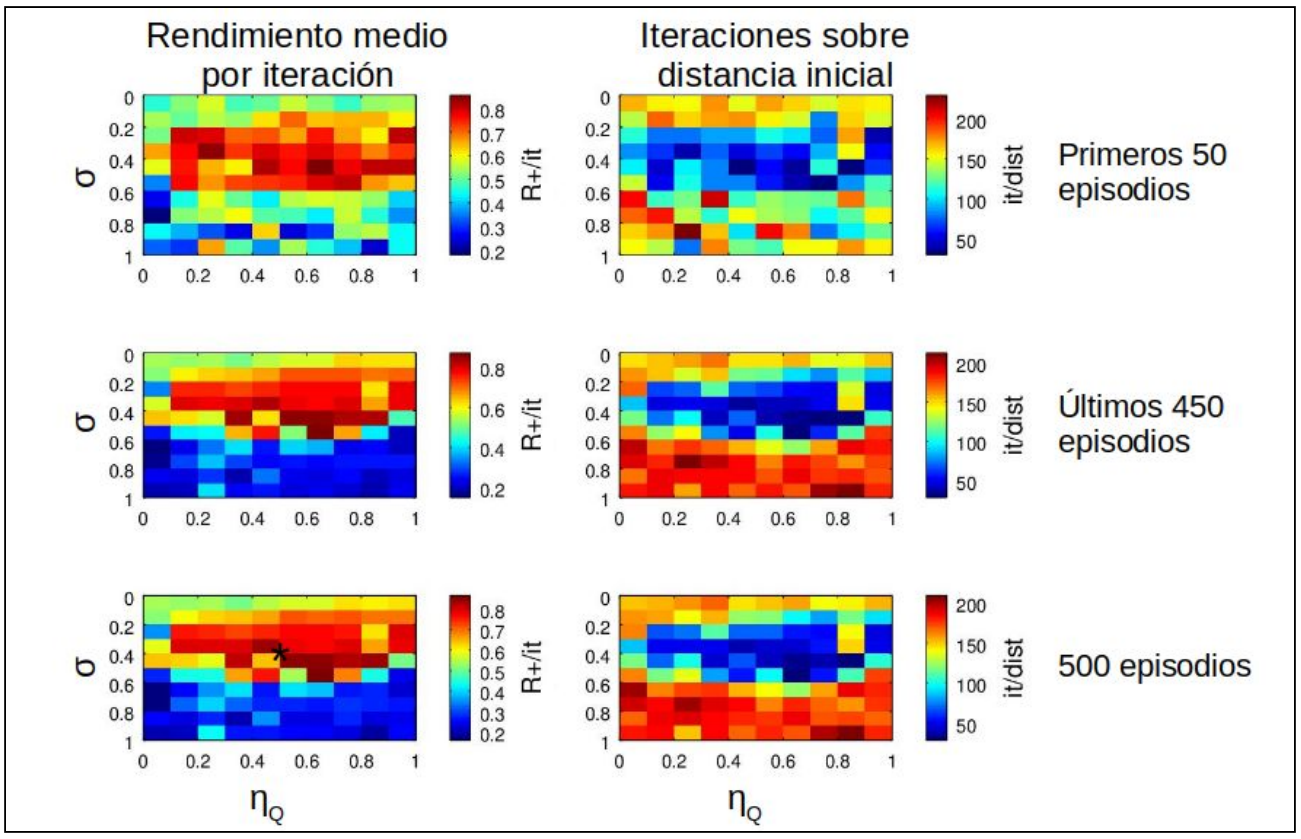

Figura 16. A la izquierda: Arriba: Rendimiento medio por iteración (R+/it) de los primeros 50 episodios para cada valor de η_Q y σ calculado; Centro: Rendimiento medio por iteración (R+/it) de los últimos 450 episodios para cada valor de η_Q y σ; Abajo: Rendimiento medio por iteración (R+/it) de los 500 episodios para cada valor de η_Q y σ. Con un asterisco se señala la posición aproximada de los parámetros que se definieron como estándar. A la derecha: Arriba: Número de iteraciones sobre distancia (it/dist) para los primeros 50 episodios al variar los valores de η_Q y σ; Centro: Número de iteraciones sobre distancia (it/dist) para los últimos 450 episodios de entrenamiento al variar los valores de η_Q y σ; Número de iteraciones sobre distancia (it/dist) para los 500 episodios de entrenamiento al variar los valores de η_Q y σ.

Sumado a esto se observó, para todos los parámetros variados, en términos cualitativos una correlación entre el cociente it/dist con el cociente R+/it, en el que, cuanto mayor fue el valor de este último menor fue el valor del primero (ver Figura 13 izquierda vs Figura 13 derecha; ver Figura 14 izquierda vs Figura 14 derecha; ver Figura 15 izquierda vs Figura 15 derecha; ver Figura 16 izquierda vs Figura 16 derecha).

5.4. Alineamiento medio en relación a la distancia

Al analizar la media del ángulo |θ| con respecto a la distancia, se puede ver que esta media decae para aquellas distancias en las que el agente se encuentra con respecto al objeto entre 1,5 y 0,6 unidades de distancia aproximadamente. En este último valor de distancia es donde se observa el mínimo valor medio de amplitud angular, apenas superior a un radian. Para aquellas distancias superiores a una unidad, rápidamente se alcanzaron valores medios superiores a 2 radianes. Para distancias menores a 0,6 unidades se observó un incremento en la media del ángulo, observándose un aumento bastante pronunciado para valores cercanos a cero (Figura 17, abajo). Algo similar ocurre con el valor mínimo al analizar solamente los episodios que fueron exitosos salvo que la curva de crecimiento del ángulo para valores más elevados que 0,6 unidades de distancia fue menos pronunciada, rondando un valor de 1,6 radianes para distancias lejanas. Para aquellas distancias menores a 0,6 unidades se vio un incremento también menos pronunciado, observándose un salto para valores muy cercanos a 0 (Figura 18, abajo). La mayor parte de las épocas registradas (N) se encontraron para valores entre 0 y 1 en ambos casos (Figura 17, arriba; Figura 18, arriba).

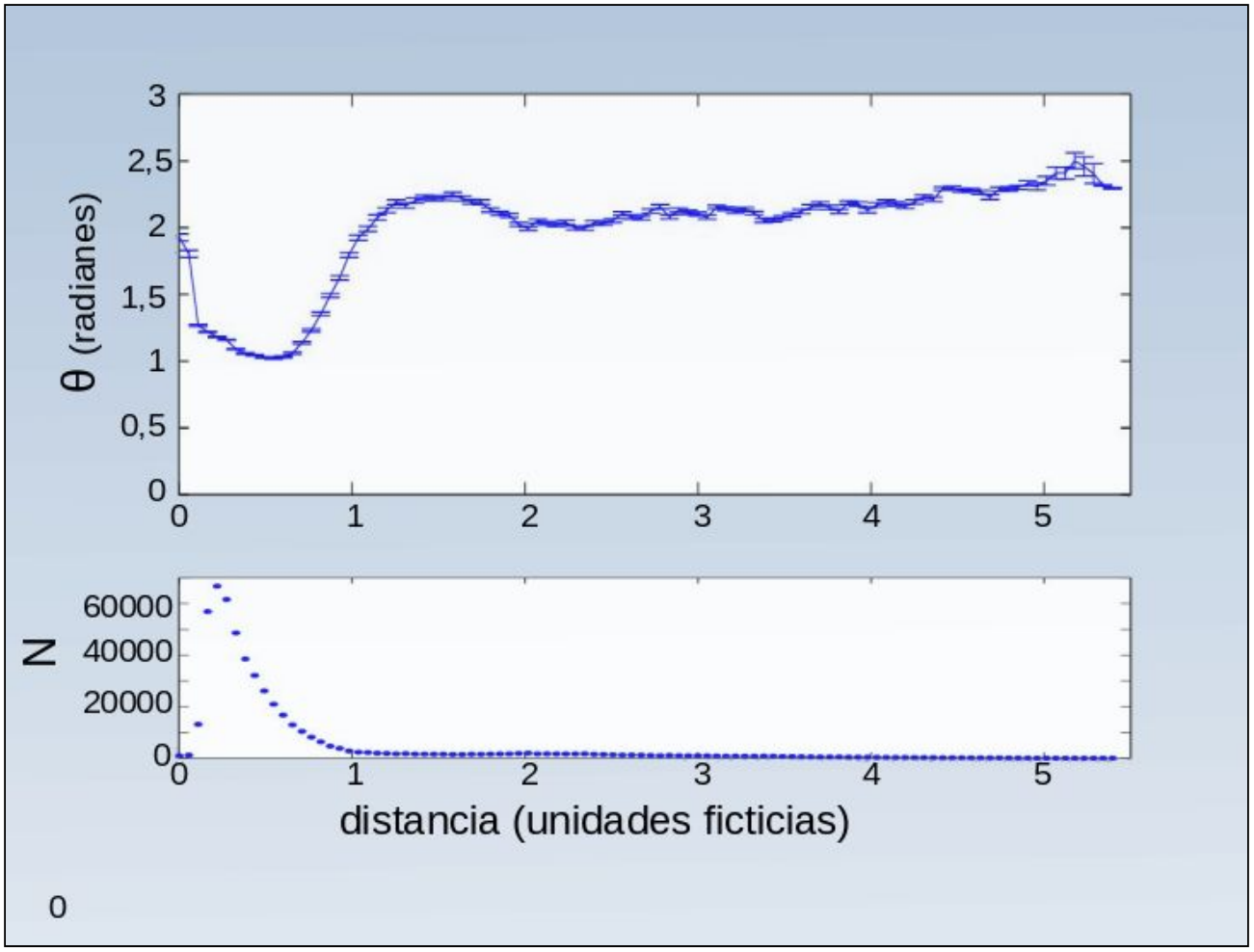

Figura 17. Arriba: Promedio del valor absoluto del ángulo θ, obtenido a partir de los 250 últimos episodios, para cada intervalo de distancia, siendo el número de intervalos de 100. Abajo: Número de datos obtenidos para cada intervalo de distancia.

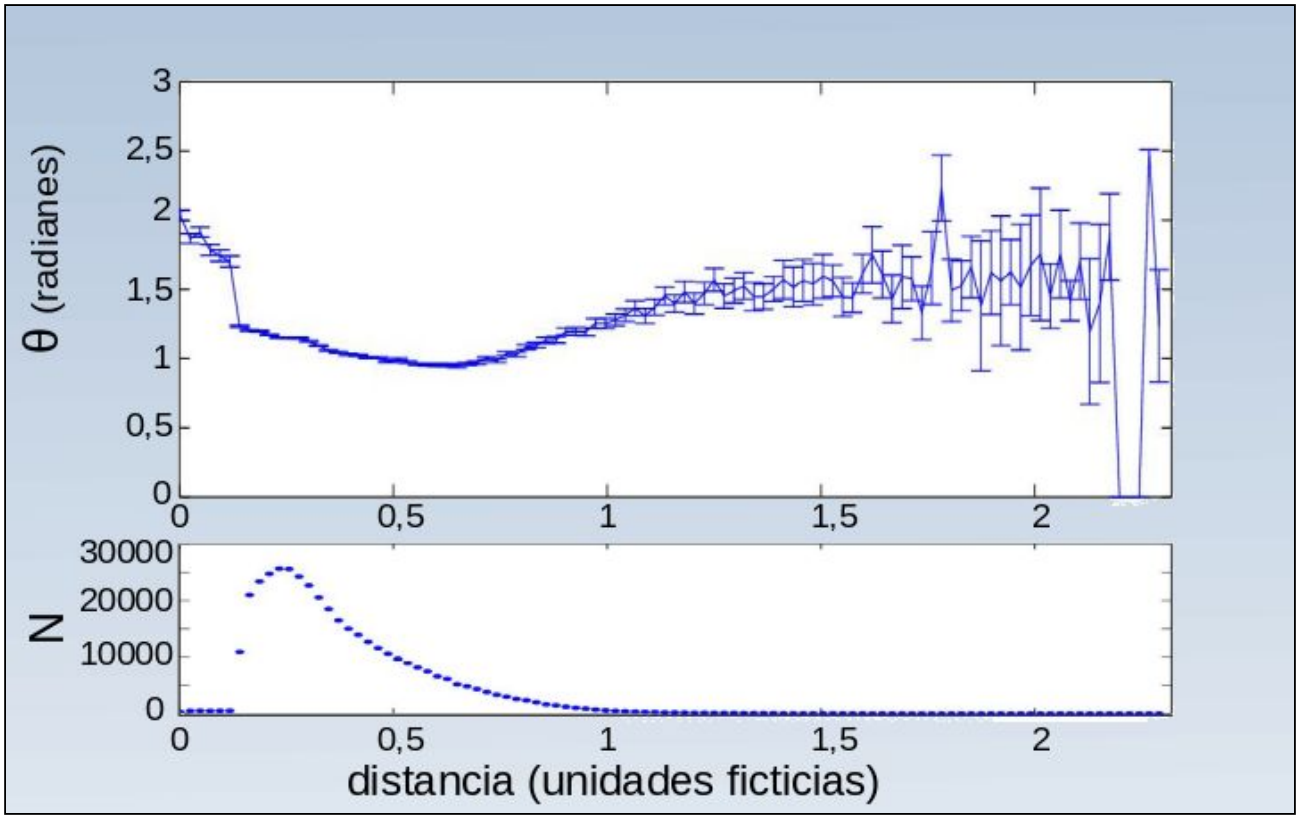

Figura 18. Arriba: Promedio del valor absoluto del ángulo θ, obtenido de los casos exitosos de los 250 últimos episodios, para cada intervalo de distancia, siendo el número de intervalos de 100. Abajo: Número de datos obtenidos para cada intervalo de distancia.

5.5. Ángulo de giro (Φ) en relación al ángulo de alineación (θ)

Los promedios de los valores del ángulo de giro Φ fueron, en general oscilantes, sobre todo para los ángulos θ comprendidos entre -π a -0,5 y 0,5 a π (Figura 19). En el primer caso el sentido de rotación fue en sentido horario mientras que para el segundo intervalo fue en sentido antihorario (lo que, como ya se mencionó en la sección anterior, se observa por el signo de Φ). Para ángulos de θ que se encontraban en el entorno de cero radianes, es decir de -0,5 a 0,5 radianes, los ángulos de rotación promedio se pueden describir como aproximadamente nulos. La pendiente fue positiva, aproximadamente, en los intervalos entre (-π y -2) y (2 y π) (radianes), siendo negativa, aproximadamente, en los intervalos que se extienden, en valores, de -2 a -0,5 radianes y de 0,5 a 2 radianes. Siendo prácticamente nula para los valores de θ comprendidos entre -0,5 a 0,5 radianes. La distribución de los datos obtendios de los ángulos θ se aproxima (visualmente) a una distribución de tipo gaussiana, estando el centro en torno al ángulo cero (siendo más precisos, presenta dos picos, aproximadamente en los 0,5 y -0,5 radianes). Por tanto, al evaluar todos estos datos, si bien no se registra exactamente una función de corrección angular de tipo sigmoide, sí parece existir en promedio una función de corrección dependiente del ángulo.

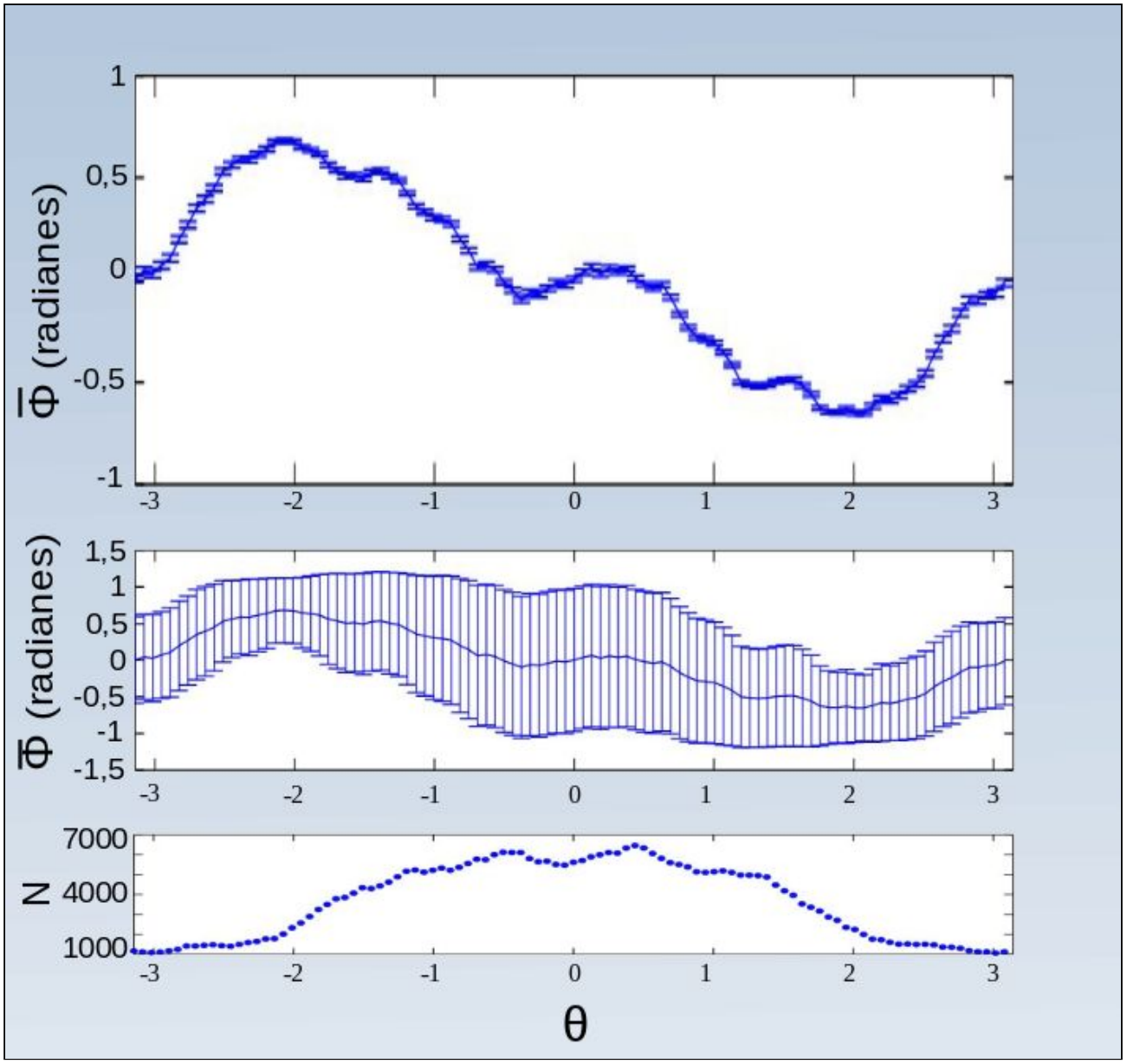

Figura 19. Arriba: Valores de Φ promediados y su respectivo error estándar en relación al ángulo θ solamente para aquellos episodios que lograron alcanzar el objetivo antes de la iteración máxima (anexo a1); el signo corresponde a si el giro realizado de una iteración a la siguiente fue horario (positivo) o antihorario (negativo). Centro: Ídem al caso anterior pero se muestra el estimativo del desvío estándar en vez del error estándar. Abajo: Distribución de los datos para cada uno de los intervalos de los ángulos θ(i).

6. Discusión

En este estudio se analizó el comportamiento de acercamiento a un objeto en el pez *Gnathonemus petersii*. Más precisamente, se trató de recrear este comportamiento suponiendo que al menos un componente del mismo era “aprendido”, implementando además una configuración simplificada. Esto para obtener de manera rudimentaria, algún concepto acerca de la lógica general de dicho patrón de navegación de este pez. Enfocándose en particular en la emergencia del alineamiento con respecto a la distancia al objeto y de las rotaciones del “pez” tendientes a alinearlo al objeto. Partiendo para ello del “aprendizaje” de aumentar la imagen eléctrica por parte del “pez” (agente). No evaluándose en este estudio la densidad de muestreo, la velocidad de avance ni la frecuencia de muestreo, las cuales podrían ser analizadas con vistas a futuro.

Para determinar si el comportamiento del agente era el esperado dentro de estas condiciones, se realizaron distintos análisis sobre los datos obtenidos en las simulaciones. Se evaluó el rendimiento al variar los parámetros, de forma de tener una idea general acerca de si el algoritmo empleado era parámetro-específico o si por el contrario tenía un amplio rango de variación permitido, lo que en última instancia daría una idea de la robustez del mismo. Sumado a esto y más importante aún, para poder determinar para qué parámetros el rendimiento era el óptimo, ya que como efectivamente se mostró el algoritmo puede variar de acuerdo con los valores de los distintos parámetros empleados por el programa. Por lo tanto, es útil saber si los valores de los parámetros que se definieron como estándar y con los cuales se desarrolló el presente estudio se aproximan a los parámetros óptimos, lo que al parecer se podría afirmar como cierto (no se puede estar seguro ya que el análisis del espacio de los parámetros fue parcial).

Alternativamente, se registró y comparó la distribución del ángulo de alineamiento con respecto al modelo “Toymodel” que como ya se mencionó había mostrado cierta similitud con los datos obtenidos a partir de muestreos experimentales. Esta comparación sirve para poder tener un estimativo acerca de cuán bien estaba el agente corrigiendo el ángulo relacionado con el alineamiento θ, así como para determinar en qué grado se recreaba a partir de este, un alineamiento dependiente de la distancia, lo que se puede traducir en cuán capaz era el modelo creado en este estudio de converger hacia el “Toymodel” y, por tanto, a una política similar a la presentada por el pez real. La presencia o no de dicha corrección angular, así como el alineamiento dependiente de la distancia, fue la que se evaluó al considerar el ángulo de giro Φ promedio con respecto al ángulo θ y el patrón de alineamiento con respecto a la distancia del pez al objeto, respectivamente.

En la presente sección se discutirán, por ende, cada uno de estos aspectos, así como las posibles causas que originan los comportamientos generales observados para este algoritmo.

6.1. Parámetros

Al estudiar el efecto de la variación de los parámetros se pretendía obtener aproximadamente los valores para los cuales el rendimiento sea más cercano al óptimo (es decir, aquel en el que se obtendría exactamente una recompensa por iteración). Para esto se realizó un estudio parcial del espacio de los parámetros. Dicho estudio, si bien excluye gran parte del espacio de las posibles combinaciones (ya que dada la cantidad de datos posibles sería excesivo el tiempo necesario a ser invertido y no es un objetivo principal de este trabajo en particular), permite tener una idea acerca de la incidencia de cada uno de los parámetros en el rendimiento del programa.

La principal observación acerca de los parámetros es que el agente tiene un buen comportamiento para los valores que se eligieron como estándar (con recompensas por iteración superiores a 0,8). Por lo que se puede afirmar que este algoritmo se corrió dentro del rango de los valores que mostraron los mejores rendimientos. Como observación secundaria, se puede destacar el hecho de que el programa en general no solía mejorar el rendimiento sino que tendía a empeorarlo para la mayoría de los parámetros variados. Por lo que (también en general) se puede decir que el algoritmo tiene un grado elevado de dependencia con los parámetros en relación al rendimiento. En particular este programa se mostró en el rendimiento:

- Dependiente de los pares: Valor-Q y valor umbral; γ y η_Q; σ y η_Q
- Poco dependiente de la variación de los pares η_a y η_s

Lo primero era esperado ya que todos esos parámetros están involucrados de algún modo con la selección de acciones y con el grado de generalización de la red. Mientras que lo segundo es interesante, por el hecho de que a priori se hubiese esperado que fuera dependiente de la constante de la velocidad de aprendizaje de los pesos de la capa oculta, pero el análisis realizado mostró lo contrario. No obstante, lo que estos parámetros posiblemente sí alteren pero que en este trabajo no se midió, es la evolución del rendimiento durante los episodios iniciales.

6.2. Rendimiento del algoritmo

En general, el programa mostró que el algoritmo para este problema en particular es dependiente de los parámetros que se elijan, a pesar de que hay ciertos parámetros que posiblemente puedan elegirse y funcionar de forma generalizable a distintos problemas. Este fue, posiblemente, el caso para los valores elegidos como estándar basándose casi en su totalidad en los valores elegidos en el artículo de Santos y Touzet (1999), en donde el rendimiento del algoritmo fue bueno, lo que

podría indicar que esos valores son bastante generalizables para distintas aplicaciones. Es interesante observar además que con estos últimos parámetros el programa tuvo un marcado mejoramiento de sus políticas durante los primeros episodios y muy rápidamente se estabilizó no mostrando mejoras más allá de las 0,85 recompensas por iteración, esto posiblemente, como se discutirá más adelante sea debido en parte a un problema de codificación de los estados. Además, otro aspecto importante, fue el hecho de que no se registraron acciones que llevaran a una recompensa nula, aumentando el tamaño de la red por medio exclusivamente de aquellas acciones que conducían a un estado en el que no se superaba el umbral de aceptación.

Por otro lado, se registró la existencia de una correlación negativa entre el número de refuerzos positivos sobre el número total de iteraciones y el número de iteraciones sobre la distancia. Esto sugiere que la función asignada al menos sirvió para explicar el componente de acercamiento del pez al objeto y una cierta tendencia a optimizar de algún modo este acercamiento cuando se obtenían refuerzos positivos.

6.3. Alineación en relación con la distancia

Para el caso de los promedios de los ángulos del valor absoluto de θ medidos (valor absoluto del ángulo que se forma entre el eje que contiene a la cabeza y la cola y el vector distancia desde el centro del pez al centro del objeto) se tratará de analizar la información obtenida para la distancia. En primer lugar, contrario a lo que se esperaba para distancias cercanas (menores a 0,6 unidades) se notó un incremento en el ángulo θ absoluto, volviéndose este abrupto para distancias muy cercanas (menores a 0,15 unidades). El incremento observado podría explicarse si suponemos que el agente posee un error que es aproximadamente independiente de la distancia a la hora de seleccionar acciones. Esto sería suficiente para explicar un incremento en la amplitud

angular promedio ya que, a medida que el pez se acerca al objeto, una misma acción supondría un incremento mayor en este ángulo que con respecto a una distancia más alejada al mismo. Además no se descarta la posible existencia de una dificultad en la codificación de los estados para estas distancias tan cercanas al objeto que tienen como consecuencia una elevada variación entre los distintos estados, lo que, en última instancia, dificultaría la selección de acciones a la hora de comparar utilizando la función de base radial. Para el intervalo de distancias entre 0 y 0,15, la acentuación en la pendiente puede explicarse dado que el objeto puede quedar "dentro" del agente con una probabilidad mayor que fuera del mismo para estos rangos de distancias, por tanto estos datos deben ser descartados por ser un artefacto del programa. Para distancias muy alejadas los valores del ángulo se ven cada vez más incrementados, esto en parte puede explicarse al comparar todos los episodios con solamente los exitosos, donde se puede observar que el mayor componente causante de ese incremento se encuentra en los casos no exitosos. Al analizar el problema de forma cualitativa, se observa que, ocasionalmente, el agente se aleja del objeto, lo que explica el elevado ángulo que se mantiene entre el eje que contiene a la cabeza y la cola y el vector distancia desde el centro del pez al centro del objeto (θ). La causa de este comportamiento de alejamiento posiblemente esté relacionada con la incapacidad del pez para codificar correctamente estados alejados al objeto, logrando superar el umbral de aceptación a la hora de seleccionar acciones a pesar de que el peso w_Q correspondiente no sea bueno, ya que este comportamiento se registró fundamentalmente cuando el pez estaba alejado del objeto.

Considerando los resultados registrados en estos casos, es difícil afirmar si hubo un alineamiento dependiente de la distancia. Sí se puede afirmar que no lo hubo para distancias menores a 0,6. El problema queda establecido para valores mayores a 0,6; donde parece existir un cierto alineamiento muy tenue a medida que se reduce la distancia del agente al objeto. Parte de este aparente alineamiento se registró debido a que el agente al alejarse tiene que tener obligatoriamente un ángulo θ mayor a π/2 (que es lo que se ve muy acentuado en el gráfico de la

Figura 17). Por esto también, al alejarse se genera una sobrerrepresentación, para distancias lejanas, de aquellos casos que no son exitosos (ya que para los casos exitosos en principio se esperaría que no se alejaran del objeto o que lo hicieran con una menor probabilidad). Al seleccionar únicamente los casos exitosos, el gráfico para las distancias más lejanas termina estabilizándose en 1,6 aproximadamente (valor cercano a π/2) lo que correspondería a un movimiento meramente azaroso. Posiblemente los valores para distancias lejanas, por tanto, se obtuvieron de forma aleatoria, lo que se traduce, analizando el algoritmo implementado, en que el pez no tomó una política aprendida para estos casos exitosos. Cuando verdaderamente tomó una política no azarosa para distancias lejanas, el pez escogió en general políticas malas (lo que se ve en la Figura 17 para aquellas distancias alejadas del objeto).

Por ende, se podría llegar a afirmar que sí se presenta algún alineamiento parcial, aunque este alineamiento registrado, dependiente de la distancia, se encuentra en un rango muy acotado y que en el gráfico posiblemente se vio acentuado a causa de problemas en la codificación de los estados para distancias relativamente lejanas.

6.4. Ángulo de giro (Φ) en relación al ángulo de alineación (θ)

Al analizar el ángulo de giro Φ con respecto al ángulo θ se puede ver que en la mayor parte del dominio el sentido de rotación (que estuvo indicado por el signo) fue el esperado, es decir, en general, el sentido de rotación promedio tendió a reducir al ángulo θ. Esto no fue lo que se registró para los ángulos θ que pertenecen al intervalo que está comprendido entre -0,5 y 0,5 aproximadamente. Para este caso el signo para el promedio que se obtuvo estuvo en torno al valor 0 aproximadamente, esto se traduce como en un no alineamiento en la amplitud angular promedio para este rango de valores de θ cuya amplitud angular es muy reducida (es decir,

cuando el pez se encuentra relativamente alineado al objeto). Esto podría explicarse por el hecho de que a pesar de que para estos casos el agente no realiza correcciones, este logra acercarse al objeto y por tanto recibir refuerzos positivos. Al aumentar el valor absoluto del ángulo θ se observa un incremento sostenido (hasta aproximadamente 2 radianes) en la corrección angular en el sentido esperado, logrando de esta forma mantener al pez en general en un ángulo θ relativamente pequeño (es decir con θ cercano a 0). Logrando de esta manera mantener al pez relativamente alineado con respecto al objeto. Esto último se puede observar al analizar la distribución de los datos generados, mostrando que para los ángulos cercanos a 0 (y especialmente cercanos a 0,5 y -0,5) el número de dichos datos para estos ángulos, en todos las simulaciones exitosas realizadas, fue significativamente mayor que para valores más alejados de este entorno. Considerando la pendiente, se pudo registrar lo que en principio se esperaba: un incremento en términos absolutos del ángulo de giro a medida que el pez se alejaba del ángulo 0, salvo para aquellos valores que se encontraron aproximadamente por encima y por debajo de 2 y -2 radianes, respectivamente, donde lo que se registró fue de forma respectiva una pendiente positiva y una pendiente negativa. Esto es debido, observando la estimación del desvío, a que se empiezan a sumar con mayor probabilidad, para valores de θ más cercanos a π (y -π) (es decir, para aquellas posiciones en las que el objeto está cada vez más cerca de la cola del pez), rotaciones en sentido horario (rotaciones en sentido antihorario para -π), haciendo que la rotación promedio tienda a anularse. Posiblemente la causa de este comportamiento sea producto de que el agente logra reducir el ángulo θ independientemente del sentido de rotación. Siendo esto último debido a que, para ángulos θ, cercanos a π o -π, el ángulo θ resultante de un movimiento horario como uno antihorario es muy similar en términos absolutos. El hecho de que el estimativo del desvío no se incremente de forma notable para estos valores del ángulo θ posiblemente se deba a que los movimientos de rotación que el pez puede realizar como máximo pueden ser, en términos absolutos, de π/3.

El estimativo del desvío mostró una variación que es relativamente elevada para ángulos θ pequeños (Figura 19) mientras que para regiones más periféricas el desvío se vio reducido. Esto

por un lado refuerza el hecho de que el agente para ángulos pequeños de θ se permite oscilaciones significativas mientras que para ángulos mayores (cercanos a los 2 radianes) se hace más necesaria la corrección angular para acercarse al objeto. Por otro lado, esto da a entender que el agente no es capaz de aumentar la alineación más allá del rango de -0,5 y 0,5, donde muestra un patrón de movimientos muy diversos.

6.5. Toymodel

El “Toymodel” es un modelo que se podría catalogar como perteneciente a la categoría de los vehículos de tipo similar a los creado por Braitenberg (1986), por presentar un accionar dado por una regla (inmutable) que a partir de un estado presenta una reacción dependiente del mismo. Tal mecanismo aplicado fue suficiente para obtener un patrón de acercamiento similar al observado de forma experimental, es decir con un alineamiento dependiente de la distancia (esto es esperable, dado que la corrección del ángulo siempre se realiza en el sentido correcto y con amplitud dependiente del alineamiento). No obstante, en el presente estudio el agente se encontró en un universo en el que no tenía “conocimiento” del ambiente en el cual está ubicado, a diferencia del “Toymodel” este pez no tiene una regla “innata” (una política no modulable) la cual le dice al agente cómo tiene que moverse dado un determinado estado (que en este caso representa la simulación de la “imagen eléctrica”). En cambio, tiene un refuerzo (determinado por la función de recompensa, que se podría decir “innato”), que recibe de acuerdo al estado que se obtiene de aplicar una determinada acción a partir de otro estado anterior (estado que contiene información relativa a la simulación de la imagen eléctrica), con el que construye una política. Dicha política en este caso es “aprendida” de forma dinámica a partir de una función de recompensa que es “innata”.

A partir de los datos discutidos en la subsección anterior se puede concluir que el algoritmo de aprendizaje implementado logró obtener un comportamiento que presenta algún tipo de corrección angular que guarda cierta similitud con el modelo creado por Gómez-Sena denominado

“Toymodel”. Es decir, se observó, en algún grado, la emergencia de una función que tiende a corregir los errores proporcionalmente con la magnitud del error. Aunque dicha similitud fue parcial, ya que debajo de un determinado umbral del ángulo θ (aproximadamente 0,5 radianes) la corrección de este ángulo se perdía y porque para valores mayores a 2 radianes el pez en promedio decrementaba su corrección en el ángulo con respecto a la corrección efectuada por el “Toymodel”. Esto puede que sea debido a que el agente posiblemente lograba recibir recompensas positivas aún manteniéndose en un rango de amplitud angular mayor a 0, lo que impedía una mayor corrección del ángulo por debajo de tal umbral del ángulo θ. Además, posiblemente por esta causa, este agente no logró en general imitar de forma similar a como hubiese sido esperado por este modelo, el alineamiento dependiente de la distancia. Mostrándose para distancias relativamente cercanas un mayor desalineamiento que con respecto a distancias más intermedias.

6.6. Aplicación del modelo en relación a aspectos biológicos

El modelo aplicado se implementó para testear si uno de los componentes observados experimentalmente durante el “comportamiento de aproximación a un objeto” (alineamiento dependiente de la distancia) podía ser explicado como un producto emergente a partir del aprendizaje indirecto (mediante la premiación del aumento de la imagen eléctrica) de otro componente de dicho comportamiento (la reducción de la distancia).

Cabe destacar que con este trabajo no se pretendió afirmar cómo es efectuado el aprendizaje en el pez real (ni si existe efectivamente un aprendizaje), puesto que:

- Este algoritmo si bien simula una red neuronal, esta red no está construida de acuerdo con las características anatómicas ni tampoco fisiológicas del pez. Precisamente porque, en principio, lo que se buscaba era explicar una propiedad emergente a partir de ciertos

supuestos básicos. Por lo que la simplificación constituyó una herramienta fundamental en el análisis de la emergencia o no de dicha propiedad.

- En el caso hipotético de que la solución hubiese mostrado un comportamiento exactamente igual al observado de forma experimental, eso igualmente no probaría que el pez esté utilizando este sistema de instrucciones para navegar. Los modelos no prueban sino que muestran la plausibilidad de ciertos supuestos y mecanismos, así como permiten formular nuevas hipótesis que se pueden testear experimentalmente.

Este modelo, sin embargo, sirvió para testear si es posible alcanzar un mecanismo razonablemente parecido al realizado por el pez (alineamiento al objeto dependiente del ángulo y acercamiento) con un un mecanismo de aprendizaje relativamente simple, que emula una red neuronal y que mapea estados a acciones, partiendo del supuesto que el agente efectúa recompensas únicamente cuando el agente se acerca al objeto (lo que se traduce en un incremento de la imagen eléctrica). Esta red a pesar de no estar basada en una red real, contiene los tres componentes de tal red, ya que hay una representación de neuronas aferentes (vector entrada), eferentes (vector salida) e interneuronas (capa oculta).

Los resultados que se registraron con este modelo sugieren que la convergencia al objeto está de algún modo (aunque de forma probabilística) relacionada con la convergencia del ángulo θ a 0. De forma probabilística debido a que es fácil suponer situaciones en las cuales el pez no se alinee (por ejemplo, en el caso que la trayectoria del pez trace una espiral en torno al objeto) y además porque las rotaciones registradas tuvieron grandes variaciones en general respecto al ángulo θ. Esto último sirve para señalar que no es suficiente recompensar el acercamiento para obtener el alineamiento relativo al objeto con respecto a dicho ángulo, aunque sí permite generar algún grado de alineamiento (al menos bajo el algoritmo aplicado y las condiciones preestablecidas).

Con respecto a la amplitud angular relativa a la distancia no se puede afirmar que ésta se redujera con la misma más allá que para un pequeño rango de distancias, observándose en algunos casos

en contra de lo que sería esperado de forma experimental un ligero incremento en tal amplitud. Esto es algo que en cierta forma empobrece un poco al modelo y por tanto plantea la necesidad de incrementar la complejidad del mismo si se quiere llegar a resultados más razonables, al menos con respecto a este alineamiento. Siendo esto debido a que en dicho caso el alineamiento respecto al objeto debía registrarse por un lado respecto al ángulo (θ) pero también debía ser respecto de la distancia tal como se observó en el “Toymodel”.

A partir de estos dos últimos párrafos, se puede afirmar que el agente tiene alguna similitud en el comportamiento en comparación con el registrado al aplicar el “Toymodel” en cuanto a la corrección dependiente del alineamiento. Aunque dicho parentesco no es suficientemente elevado como para explicar el alineamiento dependiente de la distancia observado de forma experimental y registrado también en el “Toymodel”. Por tanto se puede afirmar que, de las interrogantes planteadas en objetivos, sólo se puede afirmar parcialmente como positivo a la primera, es decir el pez presenta un corrección dependiente del alineamiento pero no significativamente dependiente de la distancia, por lo que tampoco se puede confirmar la convergencia al modelo “Toymodel”.

Otro aspecto destacable es que el pez tal vez no pueda para distancias elevadas o para distancias muy cercanas lograr una correcta codificación de los estados. A esta conclusión se llega debido a que partir de estos estados el agente tiene problemas para efectuar acciones que conlleven refuerzos positivos. Para distancias muy lejanas, posiblemente el problema surge del hecho de que todos los estados se vuelven muy similares. En cambio, para distancias muy cercanas es posible que el problema surja por lo contrario, la variación muy elevada entre estados dificultaría la correcta codificación, y por tanto, la asociación de cada uno de los estados a las correspondientes acciones. Llevando esto al pez real, la codificación se vuelve más compleja dado que la amplitud de la imagen eléctrica cae con la cuarta potencia de la distancia. Esto se traduce en que la variación entre imagen e imagen se vuelve mayor aún para distancias muy cercanas y menor para distancias muy lejanas (se acerca a 0 muy “rápidamente”).

Para darle mayor relevancia a este estudio sería de utilidad probar que el pez efectivamente implemente algún aprendizaje en algún momento de su desarrollo para aprender este comportamiento de acercamiento a un objeto y que no sea algo completamente innato (es decir, un “aprendizaje” evolutivo, ya que esto llevaría a un algoritmo con una lógica distinta, que selecciona entre poblaciones de redes neuronales). Sí se ha probado que los peces son capaces de mejorar su capacidad de navegación mediante el aprendizaje (Kieffer and Colgan, 1992; Cain, 2010; Cain et al., 2010; Schumacher et al., 2017). También se ha probado que este pez (*G. petersii*) es capaz de mejorar su comportamiento basándose en pistas alocéntricas como egocéntricas (Cain, 2010; Cain et al., 2010; Schumacher et al., 2017). Se sabe, a su vez, que es un pez de características fisiológicas y anatómicas en relación al cerebro interesantes, teniendo una muy elevada tasa de consumo de oxígeno por parte de este en relación a la masa corporal (Nilsson, 1996). Por tanto, no es imposible que exista un componente en el comportamiento de acercamiento al objeto presentado por este pez que sea aprendido.

En este estudio, el agente no se modeló para que utilizara un mapeo de tipo espacial, definiendo a un mapeo de tipo espacial como una representación métrica del ambiente en que se encuentra, de forma de obtener un registro de distintas posiciones del espacio y sus relaciones métricas relativas. El mapeo que realiza el agente es entre el estado codificado por los sensores y las posibles acciones, mediado por el refuerzo que el agente recibe. Este planteo del problema concuerda con la situación modelada en la cual el pez no tenía más referencia posible que la posición relativa al objeto y por tanto no podía utilizar ningún punto de referencia (“landmark”) para construir un mapa del espacio. Esto no significa que el pez, en otras circunstancias, no pueda utilizar puntos de referencias y construir un mapa “cognitivo”, de hecho se han realizado experimentos en los que el pez aprende a navegar a partir de puntos de referencias como sucede en general en peces (Cain, 2010; Cain et al., 2010; Schumacher et al., 2017).

Dado que no se logró una emergencia del alineamiento dependiente de la distancia (al menos del modo en que era esperado), con el fin de mejorar los resultados se podría:

- Cambiar la función de recompensa de forma de premiar únicamente aquellas transiciones que supongan un acercamiento del sensor más cefálico del pez (donde típicamente en estos peces reales se encontraría la “fóvea” electrosensorial) al objeto, dejando el incremento de la amplitud de la imagen eléctrica en todos los otros sensores sin recompensa y aplicando penalizaciones como las que se aplicaron en este programa ante disminuciones del pico de la imagen. Esto tendría posiblemente un doble efecto positivo, por un lado ayudaría al agente a alinearse y haría que el pez registre recompensas nulas (lo que es posible que afecte positivamente el rendimiento del algoritmo).
- Premiar el alineamiento del agente partiendo de un modelo similar a este debería hacer surgir de manera espontánea el comportamiento de acercamiento al objeto.
- Adaptar mejor la función con la cual se codifican y comparan los estados a este problema específico, lo que posiblemente tenga consecuencias muy positivas a la hora de mejorar el rendimiento general del algoritmo. Para lograr esto se podrían implementar amplificadores de la diferencia para posiciones alejadas al pez así como atenuadores para distancias muy cercanas de manera de que el pez pueda codificar los estados de manera más normalizable para las distintas distancias.

7. Conclusiones

En este estudio se pudo observar el efecto de la variación de los parámetros sobre el rendimiento medio del agente, así como la evolución de dicho rendimiento a lo largo del aprendizaje. En general, el “pez” mostró un rango muy acotado de valores para los distintos parámetros variados con los que el rendimiento medio tuvo una mejora a lo largo de las iteraciones. Aunque se puede afirmar que hay un conjunto de parámetros, que se tomaron en este estudio como estándar, que en general llevan a una política relativamente buena. Además el agente logró, de forma

probabilística, corregir el ángulo de nado de forma de incrementar el alineamiento aunque tal vez no en la medida esperada. Esto debido a que a pesar de la corrección angular que se evidenció, esta no fue de magnitud suficiente como para obtener un alineamiento dependiente de la distancia como hubiese sido esperado (tomando como esperado el patrón observado en el modelo "Toymodel"). Por lo que, no se puede considerar como suficiente el hecho de recompensar el acercamiento para obtener un alineamiento similar al registrado de forma experimental. Por tanto, si se desea mejorar el funcionamiento, es necesario incrementar el número de restricciones sobre el programa para lograr hacer emerger un comportamiento más apropiado.

8. Perspectivas

Dado que bajo las condiciones establecidas no se llegó al comportamiento deseado, es de carácter imperativo, si se mantiene el tipo de red, incrementar el número de condiciones que se le impone al programa con el fin de obtener un mayor parentesco con el comportamiento real del pez. Basándose en esto, como perspectiva a futuro, se podría mejorar la capacidad de alineamiento del agente al objeto cambiando la función de recompensa, de forma que, en vez de premiar aquellas iteraciones que supongan un acercamiento del pez al objeto, premiar solamente aquellas que se traduzcan en un acercamiento de la cabeza del pez al objeto. Dejando con un refuerzo nulo a aquellos otros casos que signifiquen un acercamiento (o un mantenimiento de la distancia) del pez al objeto y como negativo a los que se traduzcan en un alejamiento.

Como alternativa, sería más razonable esperar obtener de forma inversa, a través de la premiación en la reducción del ángulo θ, un acercamiento hacia el objeto en cuestión.

Sería útil, también, poder determinar si las causas que llevan al pez al patrón de aproximación a un objeto son consecuencia de un aprendizaje (aunque sea parcial) o si por el contrario es un comportamiento completamente innato del pez. No obstante, por ser esta especie actualmente irreproducible en cautiverio, puede que esta última tarea se antoje un tanto complicada.

9. Bibliografía

Bell, C.C., 1989. Sensory coding and corollary discharge effects in mormyrid electric fish. Journal of Experimental Biology 146, 229–253.

Bowdan, E., Wyse, G.A., 1996. Sensory Ecology: Introduction. The Biological Bulletin 191, 122–123. https://doi.org/10.2307/1543072

Breed, M.D., Moore, J., 2016. Neurobiology and Endocrinology for Animal Behaviorists. En: Breed, M.D., Moore, J. Animal behavior. Academic Press 1, 27-78. https://doi.org/10.1016/B978-0-12-801532-2.00002-7

Budelli, R., Caputi, A.A., 2000. The electric image in weakly electric fish. Journal of Experimental Biology 203, 481–492.

Caputi, A.A., Budelli, R., Grant, K., Bell, C.C., 1998. The electric image in weakly electric fish: Physical images of resistive objects in Gnathonemus Petersii. Journal of Experimental Biology 201, 2115–2128.

Caputi, A.A., Castelló, M.E., Aguilera, P., Trujillo-Cenóz, O., 2002. Electrolocation and electrocommunication in pulse gymnotids: signal carriers, pre-receptor mechanisms and the electrosensory mosaic. Journal of Physiology-Paris 96, 493–505. https://doi.org/10.1016/S0928-4257(03)00005-6

Caputi, A.A., Budelli, R., 2006. Peripheral electrosensory imaging by weakly electric fish. Journal of Comparative Physiology A 192, 587–600. https://doi.org/10.1007/s00359-006-0100-2

Castelló, M.E., Aguilera, P.A., Trujillo-Cenóz, O., Caputi, A.A., 2000. Pre-receptor mechanism and foveal specialisation in G. carapo. The Journal of Experimental Biology 03, 3279–3287. https://doi.org/10.1016/S0928-4257(03)00005-6

Engelmann, J., von der Emde, G., 2011. Active Electrolocation. En: Farrell A.P. processing in the Encyclopedia of Fish Physiology: From Genome to Environment 1, 375–386.

Engelmann, J., Walther, T., Grant, K., Chicca, E., Gómez-Sena, L., 2016. Modeling latency code processing in the electric sense: from the biological template to its VLSI implementation. Bioinspiration & Biomimetics 11, 055007. https://doi.org/10.1088/1748-3190/11/5/055007

Gómez-Sena, L., Budelli, R., Grant, K., Caputi, A.A., 2004. Pre-receptor profile of sensory images and primary afferent neuronal representation in the mormyrid electrosensory system. Journal of Experimental Biology 207, 2443–2453. https://doi.org/10.1242/jeb.01053

Gómez-Sena, L., Pedraja, F., Sanguinetti-Scheck, J.I., Budelli, R., 2014. Computational modeling of electric imaging in weakly electric fish: Insights for physiology, behavior and evolution. Journal of Physiology-Paris 108, 112–128. https://doi.org/10.1016/j.jphysparis.2014.08.009

Hofmann, V., Sanguinetti-Scheck, J.I., Kunzel, S., Geurten, B., Gómez-Sena, L., Engelmann, J., 2013. Sensory flow shaped by active sensing: sensorimotor strategies in electric fish. Journal of Experimental Biology 216, 2487–2500. https://doi.org/10.1242/jeb.082420

Hofmann, V., Geurten, B.R.H., Sanguinetti-Scheck, J.I., Gómez-Sena, L., Engelmann, J., 2014. Motor patterns during active electrosensory acquisition. Frontiers in Behavioral Neuroscience. 8, 186. https://doi.org/10.3389/fnbeh.2014.00186

Hofmann, V., Sanguinetti-Scheck, J.I., Gómez-Sena, L., Engelmann, J., 2017. Sensory Flow as a Basis for a Novel Distance Cue in Freely Behaving Electric Fish. Journal of Neuroscience 37, 302-312. https://doi.org/10.1523/JNEUROSCI.1361-16.2016

Hopkins, C.D., 2005. Passive Electrolocation and the Sensory Guidance of Oriented Behavior. En: Bullock, T.H., Hopkins, C.D., Popper, A.N., Fay, R.R. Electroreception. Springer Handbook of Auditory Research 21, 264–289. https://doi.org/10.1007/0-387-28275-0_10

Kandel, E.R., Barres, B.A., Hudspeth A. J., 2013. Nerve Cells, Neural Circuitry, and Behavior. En: Kandel, E.R., James, H.S., Thomas, M.J. Principles of neural science. McGraw-Hill Medical 1, 21–38

Lentz, T.L., Erulkar, S.D., 2019. Nervous system - Evolution and development of the nervous system. Encyclopaedia Britannica.

Lissmann, H.W., 1951. Continuous Electrical Signals from the Tail of a Fish, Gymnarchus niloticus Cuv. Nature 167, 201–202. https://doi.org/10.1038/167201a0

Lissmann, H.W., 1958. On the Function and Evolution of Electric Organs in Fish. Journal of Experimental Biology 35, 156–191. https://jeb.biologists.org/content/35/1/156

MacIver, M.A., 2008. Neuroethology: from morphological computation to planning. En: The Cambridge Handbook of Situated Cognition. Cambridge University Press 1, 480–504

Maler, L., 2009a. Receptive field organization across multiple electrosensory maps. I. Columnar organization and estimation of receptive field size. Journal of Comparative Neurology 516, 376–393. https://doi.org/10.1002/cne.22124

Maler, L., 2009b. Receptive field organization across multiple electrosensory maps. II. Computational analysis of the effects of receptive field size on prey localization. Journal of Comparative Neurology 516, 394–422. https://doi.org/10.1002/cne.22120

Mitchell, T.M., 1997. Reinforcement Learning. En: Mitchell, T.M. Machine Learning. McGraw-Hill Science 1, 367-388.

Northcutt, R.G., 2012. Evolution of centralized nervous systems: Two schools of evolutionary thought. Proceedings National Academy of Science of the United States of America 109, 10626–10633. https://doi.org/10.1073/pnas.1201889109

Post, N., von der Emde, G., 1999. The “novelty response” in an electric fish. Response properties and habituation. Physiology and Behavior 68, 115–128. https://doi.org/10.1016/S0031-9384(99)00153-5

Rasnow, B., 1996. The effects of simple objects on the electric field of Apteronotus. Journal of Comparative Physiology A 178, 397-411. https://doi.org/10.1007/BF00193977

Rother, D., Migliaro, A., Canetti, R., Gómez, L., Caputi, A.A., Budelli, R., 2003. Electric images of two low resistance objects in weakly electric fish. Biosystems 71, 169–177. https://doi.org/10.1016/S0303-2647(03)00124-2

Santos, J.M., Touzet, C., 1999a. Exploration tuned reinforcement function. Neurocomputing 28, 93–105. https://doi.org/10.1016/S0925-2312(98)00117-9

Santos, J.M., Touzet, C., 1999b. Dynamic Update of the Reinforcement Function During Learning. Connection Science 11, 267–289. https://doi.org/10.1080/095400999116250

Toerring, M.J., Belbenoit, P., 1979. Motor programmes and electroreception in mormyrid fish. Behavioral Ecology and Sociobiology 4, 369–379. https://doi.org/10.1007/BF00303243

Toerring, M.-J., Moller, P., 1984. Locomotor and electric displays associated with electrolocation during exploratory behavior in mormyrid fish. Behavioral Brain Research 12, 291–306. https://doi.org/10.1016/0166-4328(84)90155-4

von der Emde, G., 1992. Electrolocation of Capacitive Objects in Four Species of Pulse-type Weakly Electric Fish: II. Electric Signalling Behaviour. Ethology 92, 177–192. https://doi.org/10.1111/j.1439-0310.1992.tb00958.x

von der Emde, G., 1999. Active electrolocation of objects in weakly electric fish. Journal of Experimental Biology 202, 1205–1215.

Woergoetter, F., Porr, B., 2008. Reinforcement learning. Scholarpedia 3, 1448. https://doi.org/10.4249/scholarpedia.1448

Wyse G. A., 2013. Neurons. En: Hill, R.W., Wyse G.A., Anderson G. A. Animal Physiology 1, 295-316.

Zug, G.R., 2017. Locomotion. En: Encyclopaedia Britannica. Encyclopaedia Britannica. https://www.britannica.com/topic/locomotion

Zupanc, G.K.H., Bullock, T.H., 2005. From Electrogenesis to Electroreception: An Overview. En: Bullock T.H., Hopkins C.D., Popper A.N., Fay R.R. Electroreception, Springer Handbook of Auditory Research. Springer 21, 5–46.

10. Anexo

a1. Iteración o época número 99

a2. Se consideró no aceptable a aquellas políticas cuyo cociente de recompensas sobre iteración fue menor a 0,6