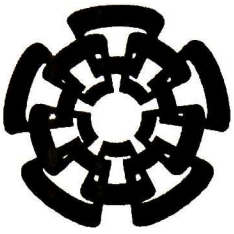




CT-699-SS1

Doc. # 2013

XX (202094.1)



Centro de Investigación y de Estudios Avanzados  
del Instituto Politécnico Nacional  
Unidad Guadalajara

# **Modelo del Sistema Cognitivo de Atención para Criaturas Virtuales**

Tesis que presenta:

**Silviano Díaz Barriga Sosa**

para obtener el grado de:

**Maestro en Ciencias**

en la especialidad de:

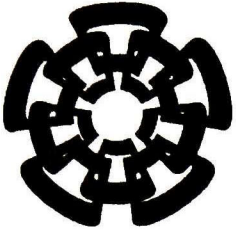
**Ingeniería Eléctrica**

Directores de Tesis

**Dr. Félix Francisco Ramos Corchado**

**Dr. Marco Antonio Ramos Corchado**





Centro de Investigación y de Estudios Avanzados  
del Instituto Politécnico Nacional  
Unidad Guadalajara

# **Modelo del Sistema Cognitivo de Atención para Criaturas Virtuales**

Tesis que presenta:

**Silviano Díaz Barriga Sosa**

para obtener el grado de:

**Maestro en Ciencias**

en la especialidad de:

**Ingeniería Eléctrica**

Directores de Tesis

**Dr. Félix Francisco Ramos Corchado**

**Dr. Marco Antonio Ramos Corchado**



CLAVIN. CT00603  
ADQUIS. CT-699-SS1  
FECHA: 22-01-2013  
PROCESO. PEN-2013  
\$ \_\_\_\_\_

# **Modelo del Sistema Cognitivo de Atención para Criaturas Virtuales**

**Tesis de Maestría en Ciencias  
Ingeniería Eléctrica**

Por:

**Silviano Díaz Barriga Sosa**  
Ingeniero en Computación  
Universidad de Guadalajara 2002-2006

Becario de CONACYT, expediente no. 243192

Directores de Tesis

**Dr. Félix Francisco Ramos Corchado**

**Dr. Marco Antonio Ramos Corchado**

# **Agradecimientos**

**A Dios y a mi familia, especialmente a mis padres, por su cariño y soporte en los momentos agradables pero sobre todo por su apoyo en los momentos más difíciles.**

**A mis amigos y compañeros de camino por haberme permitido compartir alegrías, conocimiento, estrés y demás cosas del día a día.**

**A mis asesores por su guía, consejos, recomendaciones y sugerencias en la realización de esta tesis.**

**A todas aquellas personas que directa e indirectamente me ayudaron durante estos dos años para hacer más llevadero el camino.**

**A CONACYT por proporcionarme el apoyo financiero para poder realizar los estudios de maestría.**

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Definición del Problema	1
1.2. Motivación .	2
1.3. Objetivos de la Tesis	3
1.4. Propuesta	3
1.5. Organización del Documento	5
<b>2. Otras Aproximaciones del Proceso de Atención</b>	<b>7</b>
2.1. Introducción	7
2.2. Modelos Teóricos de Atención Basados en Evidencia Neurocientífica	8
2.2.1. Modelo de Competición Sesgada	8
2.2.2. Componentes Fundamentales de la Atención	9
2.3. Modelos Computacionales de Atención	11
2.3.1. Mecanismo de Búsqueda Basado en Prominencia	12
2.3.2. Modelo Neurodinámico de Atención Visual	13
2.3.3. Modelo de Visión por Computadora para Atención Basada en Objetos y Movimiento de los Ojos	15
2.4. El Proceso de Atención en Arquitecturas Cognitivas	17
2.4.1. El Proceso de Atención en SOAR	17
2.4.2. La Atención en ACT-R	19
2.4.3. Modelo de Atención Multimodal para la Arquitectura iCub	21
2.5. Conclusiones	23



<b>3. Modelo de Atención para Criaturas Virtuales</b>	<b>25</b>
3.1. Introducción	25
3.2. Planteamiento	26
3.3. Áreas Cerebrales y Componentes Involucrados	29
3.4. Influencias <i>Bottom-Up</i> en el Proceso de Atención	33
3.4.1. Prominencia Basada en Características Físicas	34
3.4.1.1. Prominencia ICO	35
3.4.1.2. Prominencia Emocional	38
3.4.1.3. Prominencia Basada en la Forma	39
3.4.1.4. Cálculo de la Prominencia Basada en Características Físicas	40
3.4.2. Prominencia Semántica .	41
3.4.2.1. Congruencia Semántica	42
3.4.2.2. Familiaridad de los Objetos	44
3.4.2.3. Cálculo de la Prominencia Semántica	44
3.5. Influencias <i>Top-Down</i> en el Proceso de Atención	45
3.5.1. Plantillas de Atención	46
3.5.2. Relevancia Basada en Características Físicas	48
3.5.2.1. Relevancia ICO	50
3.5.2.2. Relevancia Basada en la Forma	51
3.5.2.3. Cálculo de la Relevancia Basada en Características Físicas	51
3.5.3. Relevancia Semántica	53
3.5.3.1. Cálculo de la Relevancia Semántica	53
3.6. Despliegue de la Atención	54
3.6.1. Cálculo de Prioridad y Selección Competitiva	54
3.6.2. <i>Covert Attention</i>	58
3.6.3. <i>Overt Attention</i>	58
3.6.4. IOR (Inhibition of Return)	59
<b>4. Implementación y Resultados</b>	<b>61</b>

<b>ÍNDICE GENERAL</b>	<b>III</b>
4.1. Introducción	61
4.2. Herramientas Utilizadas	62
4.2.1. iNVT	62
4.2.2. OpenCV	63
4.2.3. Alfred	63
4.3. Caso de Estudio: Inspección Libre ( <i>Free-Viewing</i> )	64
4.3.1. Resultados	66
4.4. Caso de Estudio: Búsqueda Visual ( <i>Visual Search</i> )	69
4.4.1. Resultados	70
4.5. Conclusiones .	74
<b>5. Conclusiones</b>	<b>77</b>
5.1. Conclusiones	77
5.2. Trabajo Futuro	79
<b>Referencias</b>	<b>81</b>

# Índice de tablas

3.1. Función del componente STC en el cálculo de la prominencia ICO.	37
3.2. Función del componente FEF's en el cálculo de la prominencia ICO.	37
3.3. Función del componente LIP en el cálculo de la prominencia ICO.	38
3.4. Función del componente AMY en la prominencia emocional.	39
3.5. Función del componente ETC en el cálculo de la prominencia basada en la forma.	40
3.6. Función del componente LIP en el cálculo de la prominencia basada en la forma.	41
3.7. LIP durante el cálculo de la prominencia basada en características físicas.	42
3.8. Función del componente PHC en la etapa de congruencia semántica.	43
3.9. Función del componente PHC en la etapa de familiaridad de los objetos.	44
3.10. Función del componente ITC en el cálculo de la prominencia semántica.	45
3.11. Función del componente LIP en el cálculo de la prominencia semántica.	45
3.12. Función del componente WM.	48
3.13. Función del componente PUL.	48
3.14. Función del componente STC en el cálculo de la relevancia ICO.	50
3.15. Función del componente LIP durante la etapa de relevancia ICO.	51
3.16. Función del componente ETC en el cálculo de la relevancia basada en la forma.	52
3.17. LIP durante el cálculo de la relevancia basada características físicas.	52
3.18. Función del componente ITC en el cálculo de la relevancia semántica.	54
3.19. LIP durante el cálculo de prioridad y selección competitiva.	57
3.20. Función del componente FEF en la atención abierta.	59

3.21. Función del componente SC en la atención abierta.
---

60

# Índice de figuras

2.1. Modelo de Competición Sesgada de Atención Visual. Imagen tomada de (Kastner y Ungerleider, 2000)	9
2.2. Componentes Fundamentales de la Atención. Imagen tomada de (Knudsen, 2007)	10
2.3. Modelo de búsqueda basado en prominencia. Imagen tomada de (Ittiy Koch, 2000)	13
2.4. Arquitectura general del sistema neurodinámico de atención visual. Imagen tomada de (Deco y Zihl, 2001)	14
2.5. Modelo de Atención basado en objetos y de movimiento de los ojos de Sun y colaboradores, 2008. Imagen tomada de (Sun y cols., 2008)	16
2.6. Modelo de atención NOVA en la arquitectura cognitiva SOAR. Imagen tomada de (Weismeyery Laird, 1993)	18
2.7. La arquitectura cognitiva iCub. Imagen tomada de (Vernon y cols., 2007)	21
2.8. Combinación de mapas de prominencia visual y acústico en un mapa de prominencia egocéntrico. Imagen tomada de (Ruesch y cols., 2008)	22
3.1. Diagrama general del modelo de atención propuesto en esta tesis.	27
3.2. Componentes, áreas cerebrales y sus respectivas conexiones consideradas en la propuesta de modelo de atención visual.	28
3.3. Áreas cerebrales y componentes involucrados en la atención visual (EYE, LGN, STC, ETC e ITC).	30
3.4. Áreas cerebrales involucradas en la atención visual (SC, FEF, LIP y PUL).	32
3.5. Áreas cerebrales involucradas en procesos que influyen en la atención (PHC y AMY).	33

3.6. Influencias Exógenas ( <i>bottom-up</i> ) consideradas en la solución propuesta.	34
3.7. Áreas y componentes involucrados en la etapa de prominencia basada en características físicas. Prominencia ICO (A), Prominencia Emocional (B) y Prominencia basada en la forma (C).	35
3.8. Áreas y componentes involucrados en la etapa de prominencia semántica. Congruencia Semántica (A), Familiaridad de los objetos (B).	43
3.9. Influencias Endógenas ( <i>top-down</i> ) consideradas en la solución propuesta.	47
3.10. Áreas y componentes involucrados en la etapa de relevancia basada en características físicas. Relevancia ICO (A), Relevancia basada en la forma (B).	49
3.11. Áreas y componentes involucrados en la etapa de relevancia semántica.	53
3.12. Áreas y componentes involucrados en el despliegue de la atención. (A) Atención cubierta ( <i>Covert Attention</i> ), (B) Atención abierta ( <i>Overt Attention</i> ).	55
4.1. Alfred, herramienta utilizada para mostrar de manera explícita comportamientos ligados a la atención tales como movimientos de los ojos.	64
4.2. Imágenes utilizadas para el primer tipo de inspección libre.	66
4.3. Imágenes utilizadas para el segundo tipo de inspección libre.	67
4.4. Ejecución del modelo de atención para el primer tipo de inspección libre. La imagen de entrada contiene un objeto que difiere del resto en sus características físicas, principalmente. En este ejemplo, el objeto que gana la competencia por atención (árbol de color café) se encuentra prácticamente en centro de la mirada de Alfred por lo que es atendido de manera cubierta.	68
4.5. Ejecución del modelo de atención para el segundo tipo de inspección libre. La imagen de entrada contiene un objeto que difiere del resto con respecto a su información semántica. En este ejemplo, el objeto que gana la competencia (estrella de mar de color naranja) se encuentra fuera del centro de la mirada de Alfred, por lo tanto se realiza un movimiento de los ojos para atenderlo.	68
4.6. Resultados de las simulaciones de los dos tipos de inspección visual.	69
4.7. Imágenes utilizadas en la variante <i>feature search</i> de búsqueda visual.	71
4.8. Imágenes utilizadas en la variante <i>conjunction search</i> de búsqueda visual.	71

- 4.9. Ejecución del modelo de atención para el caso de estudio de búsqueda visual en la variante *feature search*. En este ejemplo, dado que el objeto a buscar (barra horizontal roja) se encuentra ligeramente arriba del centro de la mirada de Alfred, se requirió de un movimiento ocular para atenderlo y posteriormente determinar que es el objeto que se desea encontrar. 72
- 4.10. Ejecución del modelo de atención para el caso de estudio de búsqueda visual en la variante *conjunction search*. En este ejemplo, debido a que el objeto a buscar (barra roja vertical) se encuentra prácticamente en centro de la mirada de Alfred, no fue necesario ningún movimiento de los ojos para atenderlo y determinar que ha sido encontrado. 73
- 4.11. Resultados de los experimentos del caso de estudio de búsqueda visual. 73

# Acrónimos

- 2D** Dos Dimensiones o bidimensional.
- ACT-R** Adaptive Control of Thought—Rational.
- AMY** Amygdala.
- BSD** Berkeley Software Distribution.
- ETC** Extrastriate Cortex.
- FACS** Facial Action Coding System.
- FEF** Frontal Eye Fields.
- ICO** Intensidad Color Orientación.
- IHM** Interacción Hombre-Máquina.
- iNVT** iLab Neuromorphic Vision C++ Toolkit.
- IOR** Inhibition Of Return.
- ITC** Inferotemporal Cortex.
- LGN** Lateral Geniculate Nucleus.
- LIP** Lateral Intraparietal area.
- LSF** Low Spatial Frequency.



**NOVA** Not Overt Visual Attention.

**OpenCV** Open Source Computer Vision Library.

**PEF** Parietal Eye Fields.

**PHC** Parahippocampal Cortex.

**PUL** Pulvinar nucleus.

**RGB** Red Green Blue.

**SC** Superior Colliculus.

**SOAR** State Operator And Result.

**STC** Striate Cortex.

**tIOR** temporal Inhibition Of Return.

**WM** Working Memory.

**WTA** Winner-Take-All.

# Resumen

En la actualidad ha surgido la necesidad de proporcionar a criaturas virtuales con mecanismos que les permitan exhibir comportamientos más creíbles con el fin de que puedan ser utilizadas en aplicaciones donde se requiere un mayor grado de realismo.

La atención es un proceso cognitivo de gran importancia que permite a las personas seleccionar aquella información destacada o relevante de entre toda la enorme cantidad de información que constantemente reciben a través de sus sentidos. Esta habilidad permite a los humanos concentrarse en aquella información importante para cumplir con sus metas. Acorde con la evidencia, la atención está influenciada por dos tipos principales de factores: exógenos y endógenos, los cuales determinan en un momento dado que estímulo atender.

Diversos modelos de atención tanto teóricos como computacionales han sido desarrollados, especialmente de atención visual, con el objetivo de simular esta habilidad humana. La mayoría de estos ha modelado la atención como un proceso aislado y donde la interacción de este con otros procesos cognitivos ha sido ignorada.

En este trabajo se propone un modelo de atención visual, inspirado en evidencia neurocientífica, que considera influencias exógenas y endógenas para determinar el objeto del ambiente que será atendido. El modelo propuesto se compone de diversos componentes que representan áreas cerebrales artificiales donde la atención emerge a partir de la función e interacción de estas. La influencia de información proveniente de otros procesos cognitivos ha sido contemplada, es decir, este modelo considera a la atención como un proceso que es influenciado por otros procesos cognitivos y donde esta a su vez tiene injerencia en tales procesos.

Al analizar los resultados arrojados por las simulaciones de dos casos de estudio, se comprueba que el modelo de atención que se propone representa un mecanismo mediante el cual puede proporcionarse a criaturas virtuales la habilidad de exhibir comportamientos sencillos creíbles asociados con la atención visual. Este modelo puede ser considerado como una base para generar comportamientos más complejos asociados no sólo con la atención si no también con otros procesos cognitivos.

# Abstract

Nowadays there is an increasing need to provide virtual creatures with mechanisms that allow them to exhibit more believable behaviors, so that they can be used in applications requiring a higher degree of realism.

Attention is a cognitive process of great importance that allows people to select those outstanding or relevant information from all the huge amount of information that constantly receive through their senses. This ability allows humans to focus on such important information to meet their goals. According to the evidence, attention is influenced by two main types of factors: exogenous and endogenous, which determine at any given time the stimulus to attend to.

Several theoretical and computational models of attention have been developed, especially visual attention models, in order to simulate this human ability. Most of them model attention as an isolated process and ignore its interaction with other cognitive processes.

In this work a model of visual attention, inspired by neuroscience evidence, is proposed in which exogenous and endogenous influences help to determine the object in the environment that will be attended to. The proposed model consists of several components that represent artificial brain areas where attention emerges from the interaction among them. The influence of information from other cognitive processes has been considered, i.e., this model considers attention as a process that is influenced by other cognitive processes and in turn this process also influences such processes.

By analyzing the results obtained from simulations of two case studies, we find that the proposed attention model is a mechanism by which virtual creatures can be provided with the ability to exhibit simple believable behaviors associated with visual attention. This model can be considered as a basis for generating more complex behaviors associated not only with attention but also with other cognitive processes.

# Capítulo 1

## Introducción

### 1.1. Definición del Problema

El interés por desarrollar criaturas virtuales que exhiban comportamientos más creíbles (Martinho y Ana, 2008) ha venido incrementándose de forma importante en los últimos tiempos. Los ambientes en los que las criaturas virtuales se encuentran inmersas son cada vez más complejos y con un mayor nivel de detalle. Estas criaturas han sido dotadas con diversos mecanismos que les permiten obtener información acerca de su entorno. Muchos de estos mecanismos procesan principalmente información visual del ambiente aunque algunos procesan además otro tipo de información sensorial. A través del análisis y evaluación de esta información y de los objetivos que se tengan asignados, las criaturas virtuales producen comportamientos específicos para resolver una tarea particular.

Con el fin de permitir a las criaturas virtuales exhibir comportamientos más creíbles, los mecanismos proporcionados a estas deben ser capaces de lidiar con diversos factores externos e internos. Factores externos tales como el lugar y los elementos presentes en él y factores internos tales como las expectativas, el estado emocional, el conocimiento, entre otros, deberían influir en el comportamiento de las criaturas virtuales tal y como influyen en el comportamiento de los seres humanos.

En los humanos, diferentes procesos cognitivos tales como la percepción, la memoria, las emociones, la atención, entre otros, son determinantes en la manera en que nos comportamos. Estos procesos o funciones cognitivas han sido estudiados desde varios puntos de vista y perspectivas. Disciplinas como la psicología, la neurociencia, la neurofisiología entre muchas otras continúan estudiando estos procesos cognitivos con el objetivo de determinar el cómo se llevan a cabo, sus interacciones con otros procesos y la manera en que influyen en el comportamiento. Considerando lo anterior, diversos mecanismos y técnicas se han desarrollado para simular estas habilidades cognitivas con las que cuentan los seres humanos con

el fin de dar una mayor credibilidad a los comportamientos de las criaturas virtuales. Una de estas funciones cognitivas importantes para los seres humanos es la atención. Acorde con el psicólogo del siglo XIX, William James, la atención es la toma de posesión de la mente de una forma clara y vívida de uno de los posibles objetos o pensamientos que se presentan de manera simultánea. La atención implica dejar de lado algunas cosas con el fin de tratar eficazmente a otras (James, 1890). La atención puede ser vista entonces como una especie de filtro que nos permite seleccionar de entre toda la información que los seres humanos recibimos aquella que destaca o que tiene importancia para nuestros objetivos. Al igual que otros procesos cognitivos, la atención está influenciada tanto por factores internos como externos que deben de tenerse muy en cuenta a la hora de modelar este proceso.

El proceso de atención, el cual es el tema de esta tesis, ha sido modelado tanto teóricamente como de manera computacional e incluso ha sido modelado e integrado en algunas arquitecturas cognitivas. Sin embargo, el proceso de atención, así como la interacción de este con otros procesos cognitivos, en muchas ocasiones ha sido ignorado, es decir, no se ha especificado de manera concreta o no se ha profundizado en la manera en que este se lleva a cabo aún y cuando su importancia queda de manifiesto.

Teniendo en cuenta lo previamente mencionado, es necesario el dotar a las criaturas virtuales con la capacidad de atender a aquellos elementos relevantes o destacados en el ambiente de manera similar a como las personas lo realizan. Esta habilidad le permitirá a dichas criaturas concentrarse en aspectos importantes y desechar información que no es destacada o que tiene poca o nula importancia para la tarea actual.

## 1.2. Motivación

El poder contar con criaturas virtuales capaces de mostrar comportamientos creíbles como resultado de la interacción de diversos mecanismos que simulan habilidades cognitivas, es una de las motivaciones principales para la realización de esta tesis. Criaturas virtuales dotadas con este tipo de habilidades pueden ser empleadas en muchos campos de aplicación interesantes.

Entre los campos de aplicación se incluyen los siguientes:

- Interacción Hombre-Máquina (IHM). Aplicaciones de IHM donde se requiere contar con criaturas o agentes capaces de interactuar de una manera natural y creíble con el usuario. Al contar con este tipo de entidades se mejoraría sustancialmente la experiencia del usuario ya que se tendría la impresión de estar interactuando con alguien menos mecanizado y predecible. Agentes conversacionales, tutores y asesores virtuales son solo algunos ejemplos de este tipo de aplicaciones.

- Videojuegos. En este tipo de aplicación quizá se requiera contar con personajes dotados de habilidades y capacidades mucho más complejas y realistas a las comúnmente proporcionadas.
- Aplicaciones para simular situaciones específicas. Con el objetivo de recrear situaciones donde la participación de seres humanos es riesgosa o poco viable, se pueden utilizar criaturas virtuales en su lugar. Al simular situaciones de desastre o de contingencia con este tipo de criaturas se podrían diseñar estrategias más adecuadas de evacuación, de distribución de servicios de ayuda, entre otras. Además, tales criaturas podrían emplearse también en escenarios que puedan ayudar en la capacitación de diversos profesionistas tales como personal militar, médico, etc., sin tener que utilizar personas.

Si bien es cierto que el proceso de atención representa solo una pequeña parte de la funcionalidad necesaria para que las criaturas virtuales exhiban comportamientos creíbles, no puede dejarse de lado el hecho de que es una parte fundamental ya que este proceso no solamente selecciona información relevante sino que a través de esta selección, procesos cognitivos considerados de alto nivel como la planificación y la toma de decisiones, entre otros, se ven influenciados en su funcionamiento.

### 1.3. Objetivos de la Tesis

Los objetivos de este trabajo se presentan a continuación:

- Diseñar e implementar un modelo computacional del proceso cognitivo de la atención cuyo funcionamiento esté inspirado en resultados obtenidos en la neurociencia y en otras ciencias afines. El modelo deberá de considerar factores internos y externos que influyen en la atención con el fin de poder contar con un mecanismo que semeje a aquel con el que contamos los humanos.
- Realizar el modelado del proceso de atención no como una función aislada sino como una función que interactúa con diferentes procesos cognitivos. El considerar las interacciones entre diversos procesos cognitivos permitirá exhibir un comportamiento más creíble a las criaturas que se doten con estos mecanismos.

### 1.4. Propuesta

Considerando el problema planteado anteriormente así como los objetivos de la tesis, la solución que se propone para dotar a criaturas virtuales con la capacidad de atender

estímulos relevantes en su ambiente considera mecanismos tanto de tipo *bottom-up* como de tipo *top-down*. Se les llama mecanismos *bottom-up* a aquellos mecanismos inconscientes o automáticos cuyo funcionamiento está en función de la información sensorial obtenida del ambiente. Los mecanismos *top-down* son aquellos mecanismos conscientes o voluntarios cuyo funcionamiento tiene que ver más con los objetivos, metas y conocimiento disponible. A continuación se describe de manera muy escueta los diferentes mecanismos que influyen en el modelo de atención que se propone.

La prominencia es el mecanismo *bottom-up* más importante que se considera en el modelo propuesto. Esta prominencia se calcula en base a dos factores principales: atributos físicos y congruencia semántica de los objetos presentes en el ambiente. Cada objeto contenido en el ambiente que es percibido por la criatura virtual tiene asignado un valor de prominencia.

En el modelo que se propone en esta tesis no se incluyeron muchos mecanismos de tipo *top-down*. En este trabajo se optó por elegir un tipo de factor que podría de alguna forma ser una generalización de diversos mecanismos *top-down*. Las plantillas de atención son utilizadas en este modelo para sesgar la atención hacia ciertos objetos u atributos que se correspondan con dichas plantillas. Estas plantillas podrían representar metas u objetivos de la criatura virtual. Por medio de estas plantillas se calcula la relevancia de los estímulos la cual es utilizada para sesgar la competencia por atención entre estos.

Una vez que se tiene información acerca de la prominencia de los elementos en el ambiente y de que se calcula la relevancia de los mismos en base a las plantillas de atención, esta información es combinada para determinar el elemento a atender. El modelo de atención que se propone tiene la capacidad de atender a los estímulos de dos maneras diferentes. Cabe señalar que el modelo de atención propuesto en esta tesis trabaja solo con información visual. Considerando esto, se tiene entonces que las dos maneras de atender son de manera *abierta* es decir mediante un movimiento de los ojos; la otra forma de atender es de manera *cubierta* favoreciendo el procesamiento de determinado estímulo sin necesidad de realizar un movimiento de los ojos.

En la solución propuesta se identificaron las diferentes áreas del cerebro involucradas tanto en los mecanismos *bottom-up* como en los *top-down* de atención. Enseguida se identificó la función que estas tienen y el flujo de la información en los procesos que estas llevan a cabo. El modelo fue diseñado de tal forma que la integración de otras funcionalidades para las áreas modeladas no requiera de un gran esfuerzo. Dado que la atención en conjunto con otros procesos cognitivos contribuye en la generación de comportamientos, el diseño del proceso de atención se hizo de tal forma que se permita también el modelar de una manera más eficaz la interacción entre este proceso y otros tantos con el fin de permitir comportamientos más creíbles a aquellas criaturas a las que se dote con estos mecanismos.

## 1.5. Organización del Documento

El presente documento está organizado de la siguiente manera:

- **Capítulo 2.** En este capítulo se describen los modelos de atención teóricos que inspiraron en cierta medida el modelo que se propone en esta tesis. Además, se describen algunos de los modelos computacionales de atención más relevantes en base a sus fortalezas y debilidades. Finalmente, se presenta una descripción de cómo ha sido modelado el proceso de atención en algunas arquitecturas cognitivas.
- **Capítulo 3.** En este apartado se presenta el modelo de atención propuesto como solución. Se realiza una descripción detallada de los factores y componentes que intervienen en él así como de las diferentes etapas que comprende.
- **Capítulo 4.** Este capítulo contiene información acerca de las herramientas que fueron empleadas para la implementación del modelo propuesto. También contiene los resultados que se obtuvieron en las simulaciones de los casos de estudio utilizados para evaluar dicho modelo.
- **Capítulo 5.** En el último capítulo se presentan las conclusiones obtenidas a partir del desarrollo e implementación del modelo de atención descrito en este trabajo. Finalmente se plantea el trabajo futuro que complementaría y extendería el modelo propuesto para dotarlo de mejores y mayores capacidades.



# Capítulo 2

## Otras Aproximaciones del Proceso de Atención

### 2.1. Introducción

A lo largo de los últimos tiempos se han desarrollado un número considerable de modelos de atención, principalmente modelos de atención visual. El proceso de atención ha sido ampliamente estudiado por diversas áreas y disciplinas como la neurociencia, la neurofisiología, la psicología y otras ciencias cognitivas. A partir de evidencia comportamental y biológica se han desarrollado diferentes modelos computacionales de atención para aplicaciones específicas.

A lo largo de este capítulo se realiza una descripción de algunos modelos de atención que fueron considerados de mayor relevancia. En la primera sección de este capítulo se describen dos modelos teóricos basados en neurociencia que describen el proceso de atención en base a los diferentes factores y componentes que intervienen en él. Algunos aspectos de estos modelos teóricos sirvieron de inspiración para el modelo de atención que se propone en este trabajo. En la segunda sección, se describen tres modelos que describen no sólo de manera teórica el proceso de atención sino que también plantean una solución computacional para él mismo. Estos modelos son analizados a la luz de las ventajas y desventajas que estos presentan. En la tercera sección de este apartado se hace una descripción acerca de como el proceso de atención ha sido modelado en tres arquitecturas cognitivas. De la misma manera en que se describen los modelos de la sección anterior, en esta sección se analizan también las ventajas y desventajas que ofrecen estos modelos de atención. Finalmente, en la última sección se presentan algunas conclusiones considerando los modelos de atención descritos durante este capítulo.

## 2.2. Modelos Teóricos de Atención Basados en Evidencia Neurocientífica

Como se mencionó atrás, el proceso de atención ha sido estudiado de manera exhaustiva por diferentes ciencias y disciplinas. En esta sección se presentan dos modelos teóricos de atención que están inspirados en evidencia neurocientífica. Los modelos descritos en esta sección gozan de gran aceptación por lo que se consideró adecuado tomar aspectos o ideas de estos para el desarrollo de nuestro modelo. Dado que estos modelos fueron creados en base a resultados obtenidos por la neurociencia y otras ciencias afines, se tiene información acerca de las áreas involucradas en la atención así como de la función que podrían desempeñar en ella. Esta información será de gran relevancia a la hora de diseñar nuestro modelo de atención, como se verá en el capítulo siguiente.

### 2.2.1. Modelo de Competición Sesgada

Uno de los modelos teóricos de atención visual más citados y relevantes es el modelo de competición sesgada (*biased competition*) de Desimone y Duncan (1995). Acorde con este modelo, los dos fenómenos básicos que definen el problema de la atención visual son: la capacidad limitada para procesar información y la selectividad, es decir, la habilidad de filtrar información no deseada. Los autores definen la atención como una propiedad emergente de muchos mecanismos neuronales que trabajan para resolver la competencia por procesamiento visual y el control del comportamiento.

El modelo de competición sesgada es un modelo que se centra en la vía ventral, la cual está a cargo del reconocimiento de los objetos (Ungerleider y Mishkin, 1982). Los objetos compiten en cada una de las diferentes áreas corticales que comprende esta vía, tales como la corteza visual primaria, el área V4, la corteza inferotemporal, entre otras. Esta competición está sesgada en parte por mecanismos neuronales de tipo *bottom-up* que separan las figuras de su ambiente (en espacio y tiempo) y seleccionan objetos en base a sus propiedades físicas y a la familiaridad de los mismos. La competición está sesgada también por mecanismos *top-down* que seleccionan objetos de relevancia para el comportamiento actual.

Acorde con este modelo, dentro del flujo ventral los sesgos *top-down* (plantillas) resuelven la competición principalmente entre objetos que están localizados en el mismo campo receptivo. Se dice que estos mecanismos trabajan de manera muy parecida tanto para selección espacial como para selección de objetos. La diferencia fundamental entre estos dos tipos de selección pudiera ser la fuente y la naturaleza de la plantilla de selección. Las plantillas de selección *top-down* tanto para ubicaciones como para objetos son derivadas muy probablemente de circuitos neuronales de memoria de trabajo.

En este modelo teórico de atención, los objetos actúan como un conjunto o una unidad

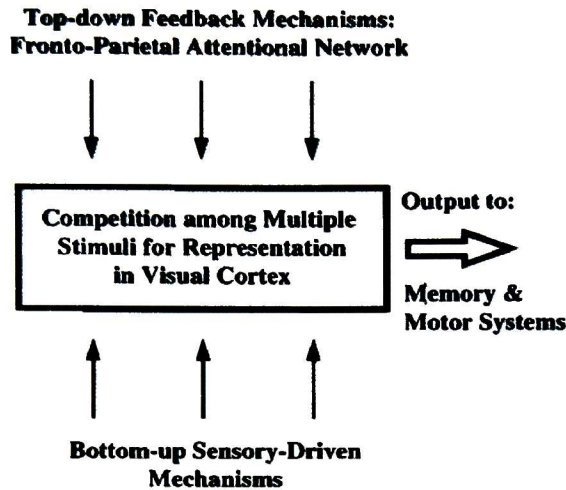


Figura 2.1: Modelo de Competición Sesgada de Atención Visual. Imagen tomada de (Kastner y Ungerleider, 2000)

en la competición neuronal. La construcción de representaciones de objetos a partir de la conjunción de diversas características aparentemente ocurre en paralelo a lo largo del campo visual antes incluso de que los objetos sean seleccionados.

A diferencia de otros modelos que consideran la atención como un foco mental de alta velocidad que escanea cada elemento en el campo visual (Eriksen y James, 1986), en este modelo la atención es vista como un conjunto de interacciones competitivas que trabaja en paralelo a lo largo del campo visual. Esta visión es importante ya que considera a la atención como un proceso rápido y dinámico.

Algunos aspectos de este modelo son importantes para el modelo que se está proponiendo, entre estos factores se incluyen el concepto de competencia entre los estímulos influenciada por los dos tipos de sesgos, así como el considerar a los objetos como unidades para la competencia. Aunque puede considerarse como un modelo antiguo, este goza de una gran aceptación y respaldo ya que ha servido como base para desarrollar otros modelos de atención tanto teóricos como computacionales.

### 2.2.2. Componentes Fundamentales de la Atención

Este modelo de atención está inspirado en el modelo de competición sesgada de Desimone y Duncan (1995). Este marco de trabajo, propuesto por Knudsen (2007), plantea la noción

de que la atención es el reflejo de las contribuciones combinadas de cuatro procesos:

- Memoria de trabajo,
- Selección competitiva,
- Control de la sensibilidad *top-down* y
- Filtrado automático de estímulos prominentes.

En la figura 2.2 se muestran de manera gráfica los componentes de la atención propuestos en este modelo. La memoria de trabajo es un tipo de memoria muy dinámica que opera sobre periodos de segundos y que almacena temporalmente información seleccionada para un mayor análisis. La selección competitiva es un proceso que determina que información obtiene el acceso a la memoria de trabajo. Por su parte, el control de la sensibilidad *top-down* es un mecanismo que regula las fuerzas de señal relativas de los diferentes canales de información que compiten por el acceso a la memoria de trabajo. Finalmente, los filtros de prominencia seleccionan estímulos que son infrecuentes en tiempo y espacio.

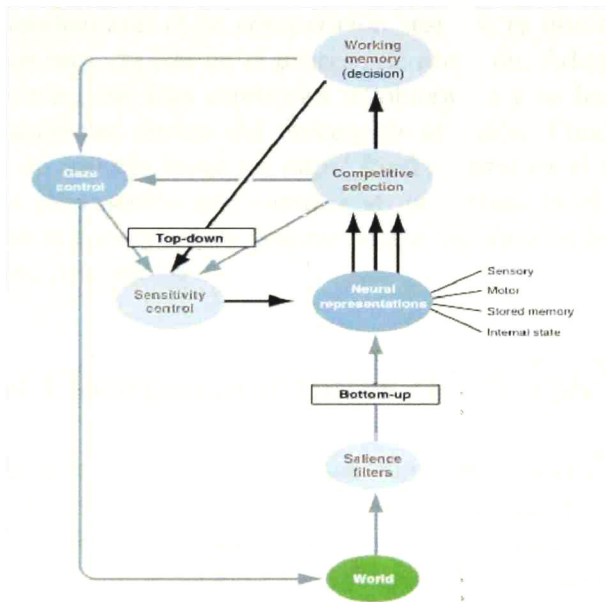


Figura 2.2: Componentes Fundamentales de la Atención. Imagen tomada de (Knudsen, 2007)

La atención selecciona la información que obtiene el acceso a la memoria de trabajo. El acceso a la memoria de trabajo es determinado por las fuerzas relativas de señal de las

representaciones de información competidoras. La fuerza de la señal es modulada de manera automática por filtros *bottom-up* de prominencia. Estos filtros seleccionan propiedades de los estímulos que probablemente son importantes. Además seleccionan estímulos de importancia biológica instintiva o aprendida. Adicionalmente, la fuerza de la señal es también modulada por señales de sesgo *top-down* que son controladas por la memoria de trabajo. El control voluntario de la atención se lleva a cabo mediante un ciclo recurrente que comprende a la memoria de trabajo, el control *top-down* de la sensibilidad y la selección competitiva.

De acuerdo con este modelo, la atención es una competencia en curso entre jerarquías de procesamiento de información que buscan el ingreso a memoria de trabajo. Las señales de sesgo *top-down* pueden de manera selectiva mejorar la representación de cierto tipo de información de tal manera que dicha información tenga una alta probabilidad de ingresar en la memoria de trabajo. Acorde con este marco de trabajo; los movimientos de los ojos junto con otros movimientos de orientación son también guiados por las decisiones tomadas en la memoria de trabajo y cooperan con las señales de sesgo *top-down* para el mejoramiento de la resolución de la información proporcionada a la memoria de trabajo.

Este modelo, de manera similar al modelo de competición sesgada, plantea la existencia de una competencia influenciada tanto por sesgos automáticos como por sesgos voluntarios. La diferencia de este modelo con el de competición sesgada es que en este se identifican y detallan cuatro procesos importantes en el proceso de atención. Además de que se identifica de una manera más precisa las áreas cerebrales involucradas y se hace una descripción del papel que podrían desempeñar dentro del proceso de atención. Finalmente, la memoria de trabajo en este marco de trabajo juega un papel fundamental en el proceso de atención ya que el fin último de la información que compete en la corteza es el ingreso a memoria de trabajo y esta misma es la encargada de generar sesgos voluntarios hacia cierta información para favorecer su ingreso a la misma.

## 2.3. Modelos Computacionales de Atención

Muchos modelos de atención computacionales han sido desarrollados a lo largo de los últimos tiempos como resultado de la evidencia que se obtiene del estudio de la atención por diferentes disciplinas. Estos modelos tratan de dar una solución desde el punto de vista computacional para el proceso de atención e identifican los factores involucrados en él así como los mecanismos necesarios para que la atención pueda llevarse a cabo. Existen dos tipos principales de modelos computacionales de atención:

- Modelos de atención basados en ubicaciones espaciales.
- Modelos de atención basados en objetos.

Existe una gran variedad de los dos tipos de modelos. La diferencia fundamental entre estos dos tipos de modelos es que en el primer tipo de modelo se atiende a ubicaciones, mientras que los modelos del segundo tipo atienden objetos. En esta sección describimos tres modelos computacionales de atención. El primer modelo es un modelo de atención basado en ubicaciones, mientras que los dos modelos restantes son modelos de atención basados en objetos. Cabe señalar que los tres modelos consideran evidencia biológica para su funcionamiento.

### 2.3.1. Mecanismo de Búsqueda Basado en Prominencia

El modelo propuesto por Itti y Koch (2000) es un modelo de búsqueda basado en prominencia para la atención visual. Este modelo está basado en el trabajo presentado por Koch y Ullman (1985), en el cual se introduce la idea de un mapa de prominencia. El mapa de prominencia es un mapa bidimensional que codifica la prominencia de los objetos en el entorno visual. La competencia entre las neuronas en este mapa da lugar a una única ubicación ganadora que se corresponde con el objeto más destacado, el cual constituye el próximo objetivo. Si esta ubicación es inhibida posteriormente, el sistema automáticamente cambia a la siguiente ubicación más destacada, dotando al proceso de búsqueda con dinámica interna.

El modelo de Itti y Koch (2000) trabaja de la siguiente manera: utilizando una imagen 2D como entrada al sistema, se extraen de esta características de bajo nivel tales como canales de color, de orientación y brillo. Para lograr esto, se le aplica a la imagen de entrada un filtro lineal obteniendo diferentes escalas espaciales de la imagen. Esta información es representada mediante mapas topográficos de características. Se consideran dos pasos fundamentales para la construcción de estos mapas. El primero de ellos consiste en cálculos de tipo *center-surround* para cada característica en diferentes escalas espaciales. El segundo paso consiste en una competencia espacial intra-característica. La información de estos mapas de características se combina de manera lineal en un único mapa que representa la prominencia local de cada una de las ubicaciones con respecto a su vecindario. El máximo de este mapa de prominencia es, por definición, la ubicación más prominente y determina la siguiente ubicación del foco de atención. Este máximo es seleccionado mediante la aplicación de un algoritmo WTA con lo cual las ubicaciones que no son las más destacadas son inhibidas para dejar activa solo la ubicación más prominente. El mapa de prominencia es dotado con una dinámica interna lo que permite al sistema perceptivo escanear la imagen de tal forma que diferentes partes son visitadas por el foco de la atención en orden decreciente de prominencia. La figura 2.3 representa una descripción gráfica del modelo explicado anteriormente.

El modelo de Itti y Koch (2000), es un modelo no muy complejo pero efectivo incluso en escenas naturales ya que es capaz de detectar de manera relativamente rápida ubicaciones u objetos destacados o prominentes. Este modelo ha servido de inspiración para muchos otros modelos de atención basados en prominencia. Un punto débil de este modelo podría ser el hecho de que el mecanismo mediante el cual se explora una imagen dada es determinista, por

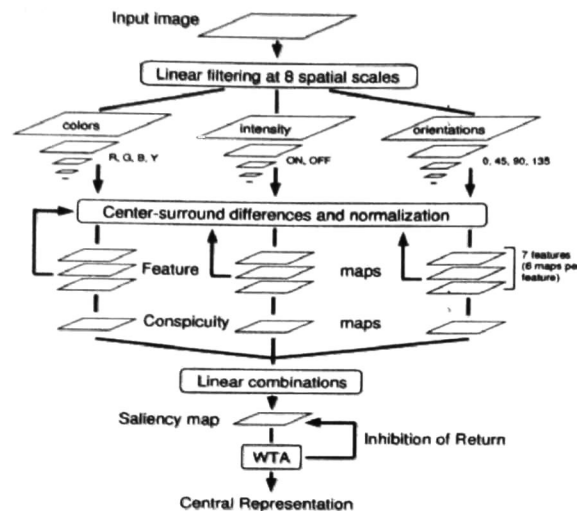


Figura 2.3: Modelo de búsqueda basado en prominencia. Imagen tomada de (Ittiy Koch, 2000)

lo que al repetir el experimento con la misma imagen se obtendría prácticamente el mismo patrón de comportamiento en caso de que no se agregue un poco de ruido al proceso o un sesgo *top-down*. Como un modelo simple de exploración, este modelo es perfecto, sin embargo si se tiene como objetivo el tener un sistema de atención en el que diferentes factores influyan en él y que interactue con otras funciones cognitivas, entonces este modelo podría quedarse corto, no obstante pudiera tomarse como base para desarrollar un sistema de atención más complejo.

### 2.3.2. Modelo Neurodinámico de Atención Visual

El modelo neurodinámico de atención visual propuesto por Deco y Zihl (2001) fue desarrollado considerando una perspectiva de neurociencia computacional. El modelo consiste de varios módulos con interconexiones de tipo *feedforward* y *feedback* las cuales describen los enlaces mutuos entre las diferentes áreas de la corteza visual (Figura 2.4). Cada módulo analiza la entrada visual de diferentes resoluciones espaciales y puede ser considerado como un vaticinador jerárquico en un nivel de resolución dado. Cada vaticinador jerárquico tiene un sub-módulo que consiste en un grupo de neuronas que llevan a cabo transformaciones de *wavelets* 2D de Gabor en un nivel de resolución determinado. Estas neuronas tienen campos receptivos que son sensibles a una región local, orientaciones particulares, ciertas simetrías

y a un nivel de frecuencia espacial particular. En este modelo se propone un mecanismo de atención selectiva basado en objetos. El control de la atención decide de manera iterativa en cuales regiones la resolución espacial debe ser mejorada. Esto se realiza mediante un mapa de prominencia que considera el nivel de la resolución de frecuencia espacial más burdo y el sesgo *top-down* proveniente de la memoria el cual es dependiente de la tarea actual.

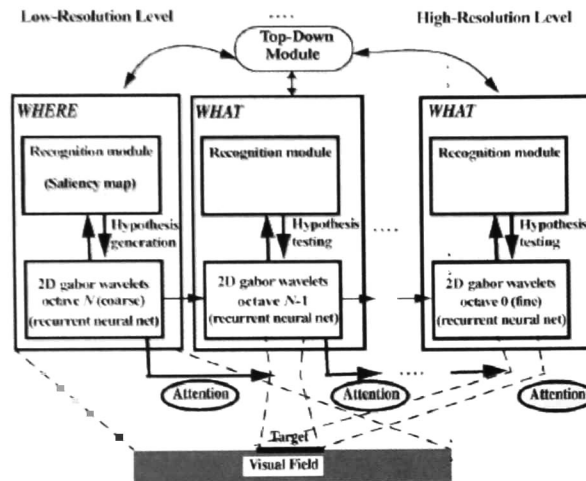


Figura 2.4: Arquitectura general del sistema neurodinámico de atención visual. Imagen tomada de (Deco y Zihl, 2001)

Al comienzo de la ejecución del modelo, sólo el módulo con el nivel más bajo de resolución está activo. La imagen es analizada en la resolución de frecuencia espacial más burda a lo largo del campo visual. Dada una tarea en particular, se genera una hipótesis basada en el módulo de reconocimiento y en la información *top-down*. Esta hipótesis consiste en la generación de un mapa de prominencia que clasifica las ubicaciones más probables en el campo visual donde el objetivo podría estar. Este módulo está relacionado con la vía dorsal (Ungerleider y Mishkin, 1982) debido a que en el nivel de resolución más burdo las neuronas en el módulo de reconocimiento contienen prácticamente sólo información acerca de la posición de los objetos pero no se tiene información detallada acerca de sus formas. El control de la atención selecciona las ubicaciones más probables donde podría encontrarse el objetivo y activa las neuronas correspondientes en el siguiente nivel de resolución. Esto se realiza mediante el control de la entrada visual en el siguiente nivel transmitiendo solo la información de la posición de los píxeles. Esta información es transmitida automáticamente por las mismas conexiones *feedback* utilizadas en las redes neuronales recurrentes basadas en *wavelets* 2D de Gabor. Está activación de los módulos continúa hasta que se tenga una clara identificación o un rechazo de la hipótesis acerca de cierto objeto. El módulo *top-down*



es activado serialmente después de la activación de cada módulo a causa de la atención. El módulo *top-down* toma una decisión acerca de la identificación o rechazo de la hipótesis en base a la tarea y la información de reconocimiento. El nivel de resolución en el cual está decisión es tomada depende de las características globales y locales del objetivo y de los distractores. Si una hipótesis es rechazada, el sistema completo comienza de nuevo ahora considerando la siguiente posición más probable indicada por el mapa de prominencia, es decir, se adopta una nueva hipótesis.

El modelo de Deco y Zihl (2001) brinda varios beneficios. Uno de estos beneficios es que está inspirado en evidencia neurocientífica lo cual es importante ya que este modelo se acerca un poco más al mecanismo de atención con el que contamos los humanos. Otra ventaja de este modelo es el de la descomposición modular del proceso de atención, es decir, representa un modelo de atención distribuido. También es importante señalar como un beneficio el hecho de que el modelo divida el procesamiento visual en las vías ventral y dorsal de esta forma el procesamiento de características visuales que resultan en el reconocimiento de los objetos se separa del proceso de ubicación espacial de los estímulos. El modelo plantea una competición influenciada tanto por factores *bottom-up* como por factores *top-down* lo cual es plausible.

El modelo descrito en esta sección no evidencia carencias de consideración. Sin embargo, en este modelo no se tiene una visión de la atención como un proceso que interactúa con otros procesos cognitivos. Si se piensa en un modelo de atención aislado limitado a ciertas aplicaciones entonces este modelo resultaría bastante apropiado. Considerando los objetivos y motivaciones establecidos en esta tesis, se tendría que idear una forma de extender este modelo con el fin de contemplar la interacción con otros procesos cognitivos.

### 2.3.3. Modelo de Visión por Computadora para Atención Basada en Objetos y Movimiento de los Ojos

El modelo que a continuación se describe fue propuesto por Sun y cols. (2008). Este modelo despliega la atención a nivel de objetos y representa un marco de trabajo biológicamente plausible para integrar cambios en la atención y movimientos de los ojos. En este modelo la atención y los movimientos de los ojos se dan en circuitos comunes subyacentes. Con el objetivo de lograr una cooperación integrada y una distinción funcional entre los cambios de atención y los cambios de la mirada, se incorporaron varios componentes en este modelo (Figura 2.5). Estos mecanismos incluyen: un sensor similar a la retina, un mapeo de prominencia de objetos visuales, inhibición de retorno temporal (que incluye memoria a corto plazo) y un interruptor dirigido por la atención.

En este modelo, los cambios en la mirada provocan que el sensor similar a la retina produzca una serie de imágenes foveadas debido al movimiento de los ojos en el tiempo. Mapas de características son obtenidos a partir de las imágenes foveadas para después formar

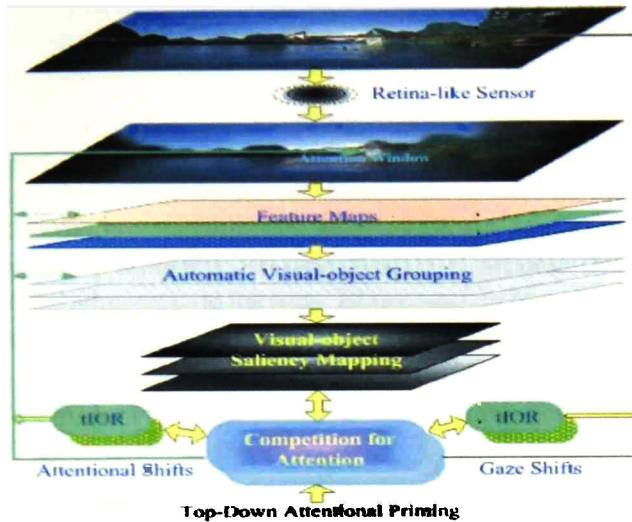


Figura 2.5: Modelo de Atención basado en objetos y de movimiento de los ojos de Sun y colaboradores, 2008. Imagen tomada de (Sun y cols., 2008)

dinámicamente los objetos visuales como una jerarquía de regiones agrupadas. Una medida de prominencia basada en color, intensidad y orientación es calculada para cada punto dentro de un agrupamiento (objeto visual). Posteriormente la prominencia de un agrupamiento es calculada como una suma de las prominencias de todos sus componentes. Después de esto los objetos compiten por atención visual. El objeto con el mayor valor de prominencia es el objeto que será atendido.

A través de un mecanismo de tipo WTA y del mecanismo tIOR el objeto que ganó la competencia captura un movimiento de los ojos para mover la fovea a su área más prominente y enseguida los cambios de atención, dentro de la ventana de atención, seleccionan objetos visuales interesantes. La posición del objetivo de un movimiento de los ojos será definida por medio de una competición entre los objetos visuales no atendidos que se encuentran fuera de la ventana de atención. El acoplamiento y desacoplamiento de la atención con los movimientos del ojo es controlado por el mecanismo interruptor dirigido por la atención.

El modelo computacional descrito en esta sección tiene como ventajas el considerar a los objetos como las unidades que compiten por el despliegue de la atención. Esto permite que este modelo pueda ser utilizado en tareas tales como la búsqueda visual en una escena determinada ya que provee una estrategia eficiente. Es decir, se requieren menos cambios de atención y movimientos de los ojos para poder localizar un determinado objeto. Otro punto positivo a considerar de este modelo es que está inspirado en evidencia biológica lo que acerca

este modelo al funcionamiento real del proceso de atención en las personas. Además, en este modelo se sugieren circuitos comunes entre dos procesos relacionados tales como la atención y el movimiento de los ojos.

Aunque este modelo representa una buena aproximación del proceso de atención, al igual que en los modelos computacionales descritos anteriormente, se dejan de lado diversos factores que podrían contribuir a enriquecer dichos modelos. Uno de esos factores es el sesgo *top-down*. Si bien el modelo contempla esta posibilidad, no se explica a detalle como esta influencia *top-down* afectaría al proceso de atención. En este modelo el proceso de atención pareciera solo tener interacción con el movimiento de los ojos. El proceso de atención interactúa con diversos procesos cognitivos tanto de bajo como de alto nivel. Sin embargo, este modelo, al igual que el anterior, no parece tener abierta la posibilidad de incluir interacciones con otros procesos cognitivos, lo cual representaría una desventaja sobre todo si se considerara la opción de dotar a criaturas virtuales con dicho modelo.

## 2.4. El Proceso de Atención en Arquitecturas Cognitivas

Hasta aquí se ha hablado acerca de diferentes modelos teóricos y computacionales del proceso de atención. Ahora es tiempo de ver como ha sido modelado dicho proceso dentro de las arquitecturas cognitivas, es decir, en una arquitectura donde diversos sistemas que representan procesos cognitivos se llevan a cabo. Estas arquitecturas son proporcionadas a agentes para brindarles la posibilidad de comportarse de una forma más inteligente y de manera autónoma. Aunque existen muchas arquitecturas cognitivas, muchas de estas pasan por alto el proceso de atención. En esta sección discutiremos como la atención ha sido modelada en tres diferentes arquitecturas cognitivas. Si bien no son las únicas en modelar el proceso de atención, cabe señalar que la forma en que estas lo modelan destaca de entre las demás. En esta sección se hace un análisis de estos modelos considerando los puntos a favor y en contra que cada uno de estos tiene.

### 2.4.1. El Proceso de Atención en SOAR

SOAR es una arquitectura cognitiva de resolución de problemas para inteligencia artificial. El diseño de SOAR está basado en la hipótesis de que todo comportamiento deliberativo orientado a metas puede ser formulado como la selección y aplicación de operadores a un estado. Un estado es una representación de la situación actual de la resolución de un problema; un operador transforma un estado y una meta es un resultado deseado de la resolución de un problema (Laird y Congdon, 2012).

SOAR no considera la atención como un proceso separado del resto de la cognición sino como una etiqueta para un conjunto de fenómenos que emergen cuando un sistema con la capacidad de un alto grado de paralelismo tiene que generar comportamiento serial. Siempre que SOAR tenga que tratar con un objeto o con un conjunto reducido de objetos de entre una cantidad más grande que podría tratar, entonces se dice que Soar esta atendiendo a ese objeto o ese pequeño conjunto de objetos. En SOAR la atención está ligada al cuello de botella serial impuesto por la selección de un solo operador a la vez (Young y Lewis, 1999).

El modelo de SOAR que de una manera más explícita trata con el proceso de atención es el de Weismeyer y Laird (1993). En su trabajo, Weismeyer y Laird (1993) proponen un modelo de atención cubierta (*covert attention*), NOVA. Este modelo se ilustra en la figura 2.3 y es considerado como el primer modelo de atención cubierta basado en operadores. Cada componente de NOVA opera de forma semi-autónoma a una tasa particular conocida como su ciclo de tiempo. Por ejemplo, el ciclo de tiempo de percepción es de 100ms, definido como el tiempo que le lleva a un estímulo viajar desde la la retina hasta el almacenamiento de imagenes visuales, el cual forma parte de la memoria de trabajo. NOVA considera dos operadores visuales en su funcionamiento: ATTEND y RECOGNIZE. Además de que considera dependencias perceptuales para los cambios de atención cubierta. Una dependencia perceptual se da cuando la creación de un operador que es necesario para una tarea depende de un estímulo visual que aún no ha sido llevado a la memoria de trabajo.

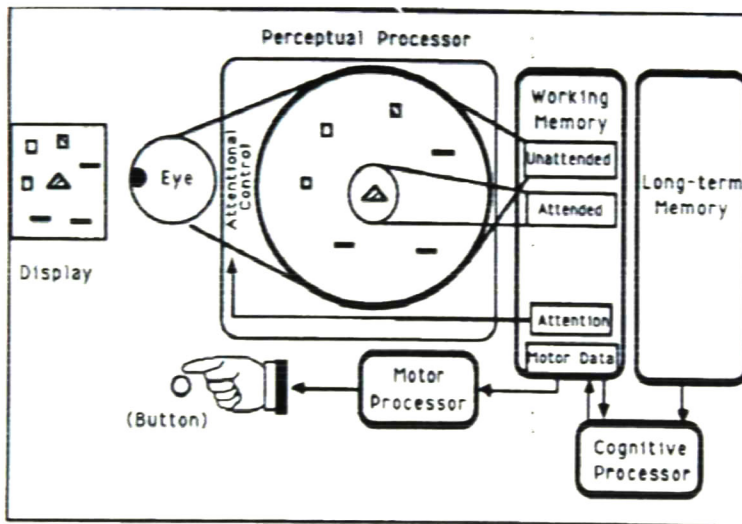


Figura 2.6: Modelo de atención NOVA en la arquitectura cognitiva SOAR. Imagen tomada de (Weismeyer y Laird, 1993)

En NOVA la atención visual debe de preceder al proceso de identificación y ambos, aten-

ción e identificación, son acciones deliberativas que son llevadas a cabo mediante los operadores ATTEND y RECOGNIZE. El lente de la atención es una simple región ovalada que recorre el campo visual separándolo en dos regiones: región atendida y regiones no atendidas. Las características atendidas son utilizadas en el proceso de reconocimiento mientras que las no atendidas son utilizadas como señal para un cambio de atención. Sólo después de que las características atendidas son transformadas en símbolos mediante el operador RECOGNIZE, las identidades de los eventos visuales pueden ser reportadas o influir en el comportamiento.

Hablando acerca de los beneficios que ofrece el modelo de atención NOVA pudiera considerarse como una ventaja el hecho de considerar la atención no como un mecanismo o componente aislado sino como parte del proceso cognitivo de la arquitectura. También cabe destacar que se tiene la noción de cambio de atención en base a los necesidades o dependencias que se tengan. Otro de los beneficios que ofrece este modelo es el hecho de contar con un ciclo de operación definido y cuantificado temporalmente, es decir, se sabe la duración y latencia de cada una de las etapas del modelo. Estas duraciones son relativamente cortas por lo que se puede decir que el modelo opera de manera rápida.

Con respecto a las desventajas de este modelo, podemos considerar su poco o nulo concepto de sesgo automático o *bottom-up* para influir en el proceso de atención. En otras palabras el proceso de atención se reduce a mecanismos voluntarios o de tipo *top-down*. Quizá para algunas aplicaciones esto no represente un gran inconveniente, pero para nuestro propósito, el cual es el de dar mayor credibilidad a las criaturas virtuales este modelo podría quedarse corto ya que es bien sabido que en la atención no solo influyen aspectos voluntarios sino también factores inconscientes dependientes de la naturaleza física de los estímulos.

### 2.4.2. La Atención en ACT-R

ACT-R es una arquitectura cognitiva cuyas construcciones reflejan afirmaciones acerca de la cognición humana la cual está basada en numerosos hechos derivados de experimentos psicológicos (Anderson y cols., 2004). ACT-R es considerada como una arquitectura cognitiva híbrida ya que contiene tanto elementos simbólicos como sub-simbólicos. Su estructura simbólica es un sistema de producción y su estructura sub-simbólica es representada por un conjunto de procesos paralelos que pueden resumirse en un número de ecuaciones matemáticas. Las ecuaciones sub-simbólicas controlan muchos de los procesos simbólicos y además son responsables de la mayoría de los procesos de aprendizaje en ACT-R.

Los principales componentes de la arquitectura ACT-R son: módulos, buffers y el *pattern matcher*. Existen dos tipos de módulos:

- Módulos perceptuales-motores, los cuales se encargan de la interfaz con el mundo real. Los módulos perceptuales-motores más desarrollados en ACT-R son el módulo visual y el manual.

- Módulos de memoria. existen dos tipos de módulos de memoria en ACT-R: memoria declarativa, que consiste de hechos; y memoria de procedimiento, la cual está conformada por producciones. Las producciones representan el conocimiento acerca de cómo hacemos las cosas.

ACT-R obtiene el acceso a sus módulos (excepto para el módulo de memoria de procedimiento) a través de buffers. El contenido de los buffers en cualquier momento representa el estado de ACT-R en ese instante. El *pattern matcher* busca una producción que se corresponda con el estado actual de los buffers. Sólo una de tales producciones puede ser ejecutada en un momento dado. Esa producción, cuando se ejecuta, puede modificar los buffers y de esta forma cambiar el estado del sistema. Así pues en ACT-R la cognición se despliega como el disparo de una sucesión de producciones.

En el trabajo de Anderson y cols. (1995) se extiende la arquitectura cognitiva ACT-R mediante la inclusión de una teoría de atención visual. En esta teoría las reglas de producción son capaces de dirigir la atención hacia características visuales primitivas en el arreglo visual. Cuando la atención es dirigida hacia una region particular, las características en dicha región pueden ser sintetizadas en *chunks* declarativos en la memoria de trabajo de ACT-R para que puedan ser sujetos a un procesamiento de mayor nivel. Cabe señalar que en este modelo no se diseñó una nueva teoría de atención visual sino que se adaptó y se embebió en ACT-R una teoría ya existente. La teoría implementada fue la teoría del foco de atención visual en la que un foco de tamaño variable puede ser movido a lo largo y ancho del campo visual.

En el modelo de atención visual de ACT-R se asume que antes del reconocimiento de un objeto sus características están disponibles como parte del objeto pero no como el objeto mismo. Con el fin de que la teoría de alto nivel de ACT-R sepa que hay en su ambiente, debe de moverse el foco de atención sobre el campo visual. En ACT-R las llamadas para los cambios de atención son controladas por disparos explícitos de reglas de producción. De ahí que se requiera de más tiempo para codificar la información visual.

Algunos aspectos positivos que se rescatan del modelo de atención visual propuesto en ACT-R es que modela de manera efectiva la limitación para procesar información que tenemos los seres humanos. Esto es importante ya que si se desea tener criaturas con comportamientos más parecidos a los humanos este modelo podría cumplir en cierta medida con este requisito. Otro punto a destacar es el hecho de que este modelo fue utilizado exitosamente en experimentos interesantes tales como la resolución de ecuaciones, con lo cual queda de manifiesto su utilidad.

Como todo modelo, el modelo de atención visual de ACT-R tiene debilidades. La principal debilidad desde nuestro punto de vista es que al igual que el modelo de atención de SOAR no se consideran aspectos automáticos en el proceso de atención. Sin duda esto representa un obstáculo si se tiene en mente proporcionar a criaturas virtuales este modelo ya que deja de lado aspectos que podrían hacer que dichas criaturas se comportaran de una manera

más creíble. Finalmente, en este modelo no se plantea una competencia explícita entre los estímulos u objetos percibidos del ambiente dado que las reglas de producción para dirigir la atención están sesgadas de antemano a ciertos estímulos o atributos.

### 2.4.3. Modelo de Atención Multimodal para la Arquitectura iCub

La arquitectura iCub (Vernon y cols., 2007) es una arquitectura cognitiva que fue desarrollada para integrarse al humanoide iCub. La arquitectura iCub comprende una red de circuitos distribuidos multifuncionales perceptuo-motores que cooperan y compiten entre sí, un circuito de modulación que lleva a cabo la selección de la acción mediante la desinhibición de los circuitos perceptuo-motores y un sistema que realiza anticipaciones a través de simulaciones de tipo percepción-acción. El circuito de modulación comprende tres componentes: memoria auto-asociativa, selección de acción basada en campos neuronales dinámicos y motivación.

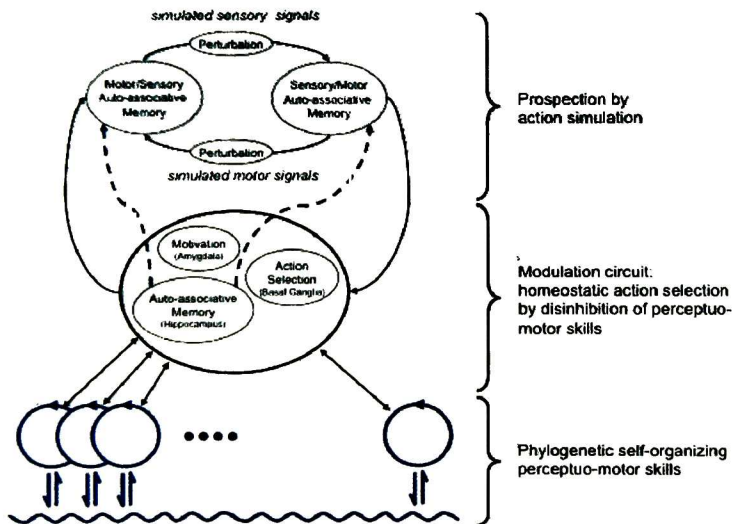


Figura 2.7: La arquitectura cognitiva iCub. Imagen tomada de (Vernon y cols., 2007)

Respecto al proceso de atención, se diseñó un modelo de atención de tipo *bottom-up* multimodal basado en prominencia para el robot humanoide iCub (Ruesch y cols., 2008). En este modelo las decisiones acerca de hacia donde mover los ojos y el cuello del humanoide

están basadas en mapas de prominencia acústicos y visuales. Estos mapas de prominencia son combinados en un marco de referencia egocéntrico.

La información de prominencia visual es calculada mediante la combinación de los resultados obtenidos de diversos sub-filtros en un mapa visual final. Estos sub-filtros calculan la prominencia visual en base a cuatro características visuales: intensidad, color, dirección y movimiento. Con respecto a la prominencia auditiva, esta se calcula en base al grado de elevación del sonido. La posición de una fuente de sonido se calcula utilizando diferencias de tiempo interaural. Una vez que se tiene los mapas de prominencia visual y acústico se convierten a un marco de referencia egocéntrico expresado en coordenadas esféricas (azimut y elevación). Una vez que la información acerca de la prominencia se convirtió en un marco de referencia común, se procede entonces a combinar esta información para tener un sólo mapa final (Figura 2.8). Por medio de este mapa final el robot humanoide decide que ubicación atender, para esto, quizá sea necesario mover el cuello y los ojos. El proceso de atención selecciona un punto correspondiente a la ubicación que tiene asociado el valor de prominencia global más alto.

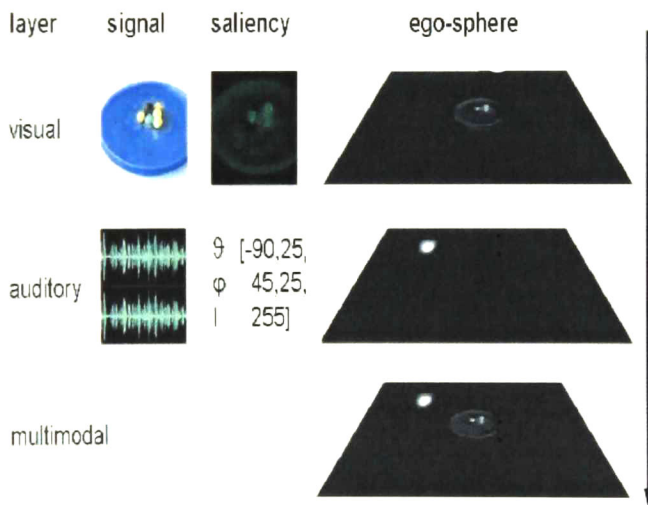


Figura 2.8: Combinación de mapas de prominencia visual y acústico en un mapa de prominencia egocéntrico. Imagen tomada de (Ruesch y cols., 2008)

En este modelo se diseñó un mecanismo para la inhibición del retorno (IOR). Mediante este mecanismo se evita que ubicaciones que ya han sido previamente atendidas vuelvan de nueva cuenta a ser atendidas por el robot. Este mecanismo atenúa la prominencia de las



ubicaciones cercanas al foco de atención por una cierta cantidad de tiempo.

El modelo de atención descrito en esta sección tiene varios argumentos a favor y otros tantos en contra. Lo positivo de este modelo es que no sólo considera información sensorial de una modalidad sino de dos, lo cual enriquece el modelo ya que no sólo se plantea una competencia de los estímulos intra-sensorial sino que además se contempla una competencia inter-sensorial. El hecho de utilizar este modelo en un humanoide es de gran relevancia ya que se demuestra su utilidad en aplicaciones reales. Otro punto a resaltar de este modelo, es que se proporciona al humanoide iCub un comportamiento de exploración el cual es el resultado de la interacción entre el proceso de atención selectiva y los mecanismos de inhibición.

No podemos dejar de mencionar ciertos aspectos que consideramos como un tipo de debilidad de este modelo de atención. La más notoria es que este modelo no está influenciado por mecanismos *top-down* los cuales son de gran importancia para el proceso de atención selectiva. Quizá en una etapa posterior se consideren estos factores, pero por el momento no sería conveniente utilizar este modelo si queremos modelar la atención de manera similar a como se da en los seres humanos. Como este marco de trabajo es un modelo de atención espacial se deja de lado aspectos importantes que influyen en el proceso de atención tales como la información semántica y la familiaridad de los objetos en la escena, por mencionar algunos.

## 2.5. Conclusiones

Después de analizar algunas aproximaciones del proceso cognitivo de atención, podemos concluir que la mayoría de estas considera a la atención como una competencia entre los estímulos por procesamiento visual. Los modelos teóricos analizados en este capítulo plantean la competencia a nivel neuronal y consideran a la memoria de trabajo como el destino o el lugar por el cual los diferentes elementos en el ambiente compiten para ingresar. Ambos modelos consideran factores inconscientes y conscientes que influyen en el proceso de atención. Dado que estos modelos tienen gran aceptación en la comunidad científica, sería deseable considerar aspectos de estas aproximaciones para el diseño de un modelo computacional de atención con respaldo teórico.

Los modelos computacionales analizados, en general, modelan la atención como una propiedad emergente de mecanismos *bottom-up* y *top-down*. Estas aproximaciones computacionales modelan los factores *bottom-up* de una manera detallada. En general se contemplan propiedades físicas de bajo nivel tales como color, orientación, etc. Los modelos analizados de este tipo, ven una clara relación entre la atención visual y el movimiento de los ojos. En general, en las aproximaciones computacionales de la atención, no se contemplan interacciones o influencias de esta con otras funciones cognitivas tales como la memoria, las emociones, entre otras. En el cerebro humano, varias funciones cognitivas interactúan para generar com-

portamientos específicos. Aislar el modelo de la atención o no considerar influencias de otros procesos cognitivos podría resultar en una aproximación incompleta para los objetivos que se persiguen en esta tesis.

Como se mencionó en el análisis del proceso de atención en las arquitecturas cognitivas, la mayoría de estas consideran a la atención como un proceso influenciado mayormente por aspectos voluntarios o dependientes de la tarea actual. Sin embargo, la arquitectura iCub si considera aspectos como los atributos físicos de los estímulos tanto visuales como auditivos. Al igual que en las aproximaciones computacionales analizadas, estos modelos no contemplan influencias en la atención derivadas de información proveniente de otros procesos cognitivos concurrentes. En muchas de las arquitecturas cognitivas, el proceso de atención se da por sentado al igual que otros procesos cognitivos considerados de bajo nivel. Sin embargo, el proceso de atención tiene gran influencia en el comportamiento ya que al actuar como un filtro de la información que se obtiene del ambiente, la atención de cierta forma decide que información utilizar para guiar el comportamiento. Entonces, si se requiere que las criaturas virtuales exhiban comportamientos creíbles será necesario dotarlas con un sistema de atención que les permita enfocarse en estímulos de manera similar a como el proceso de atención en las personas lo hace. Sin duda esto representa un enorme reto ya que el funcionamiento del cerebro es aún un misterio, no obstante con la evidencia disponible podría crearse un modelo de atención capaz de mostrar si no comportamientos idénticos a los de los humanos si comportamientos más creíbles.

# Capítulo 3

## Modelo de Atención para Criaturas Virtuales

### 3.1. Introducción

A lo largo de este capítulo se presenta nuestra propuesta para modelar el proceso de atención visual. Tomando en cuenta las fortalezas y debilidades de las aproximaciones analizadas en el capítulo anterior así como los objetivos de esta tesis, se propone un modelo de atención inspirado en evidencia neurocientífica el cual es influenciado por factores exógenos y endógenos, es decir, factores *bottom-up* y *top-down*, respectivamente. Cabe señalar que el modelo que se propone contempla influencias e información proveniente de otros procesos cognitivos tales como la memoria y las emociones con el fin último de tener un mecanismo de atención que pueda ser dotado a criaturas virtuales para que puedan exhibir comportamientos más creíbles.

La organización del capítulo comprende seis secciones principales. En la segunda sección se hace un planteamiento general del modelo de atención que se propone. La tercera sección enumera los diferentes componentes y áreas cerebrales considerados en esta aproximación que se encuentran involucrados en la atención. La descripción de los mecanismos *bottom-up* que influyen en la atención y que fueron considerados en esta aproximación son descritos en la cuarta sección. Las influencias *top-down* consideradas en este modelo se abordan en la quinta sección. Por último, se detalla la manera en que la información de los dos tipos de influencias se combina para el despliegue de la atención.

## 3.2. Planteamiento

Como se mencionó anteriormente, el proceso de atención ha sido ampliamente estudiado y modelado desde distintas perspectivas. Continuamente se obtiene nueva evidencia acerca de como la atención se da en los seres humanos, sin embargo aún se desconoce la forma precisa de operar de ciertos mecanismos y el papel exacto que algunas áreas cerebrales desempeñan en este proceso.

Diversas teorías han surgido que tratan de explicar la manera en que la atención opera. Muchas de estas consideran dos etapas principales. En la primer etapa (*pre-attentive*), la atención está distribuida de manera uniforme en toda la escena visual y el procesamiento de la información se lleva a cabo en paralelo. En la segunda etapa (*attentive*), la atención esta concentrada en un área específica (foco de atención) de la escena visual y el procesamiento se da de manera serial (Eriksen y James, 1986). Sin embargo, existen otras corrientes que difieren de la anterior. Una de estas considera a la atención como una propiedad emergente de la competencia entre los estímulos por procesamiento (Desimone y Duncan, 1995). Existe aún controversia entre las dos corrientes ya que existe evidencia que respalda ambas teorías. Dado que nuestra intención no es detallar de manera precisa como se da la atención en los seres humanos sino más bien generar un modelo basado en evidencia biológica que nos permita replicar comportamientos similares a los de los humanos, se decidió tomar aspectos de ambas corrientes.

El modelo que se propone en este trabajo (Figura 3.1) es un modelo de atención visual basado en objetos, es decir, los objetos son los que compiten por procesamiento de alto nivel y por el ingreso a memoria de trabajo. La entrada del sistema de atención propuesto son imágenes 2D, las cuales representarían el ambiente percibido por la criatura virtual en un momento dado. El modelo está centrado en la vía ventral visual, la cual acorde con Ungerleider y Mishkin (1982), se encarga del reconocimiento de los objetos y donde diferentes procesos relacionados con el proceso de atención se llevan a cabo. Algunos de estos procesos son el procesamiento de los propiedades básicas de los componentes, el agrupamiento de estas propiedades básicas en representaciones visuales y la extracción de información semántica de las representaciones visuales, entre otros (VanRullen, 2003).

Con el fin de modelar la atención de manera similar a como ocurre en las personas, se contempló influencias exógenas y endógenas en el modelo de atención. En esta aproximación se consideraron influencias exógenas tales como la prominencia de los estímulos. La prominencia (*saliency*) ha sido definida desde un punto de vista psicológico y computacional como la distinción física de un objeto. La prominencia es una propiedad relativa que depende de la relación de un objeto con respecto a otros objetos en la escena (Fecteau y Munoz, 2006). Existen muchas propiedades físicas diferentes que pueden hacer a un objeto más prominente que otros en la escena visual tales como la intensidad, el color, la orientación, la forma, el movimiento, entre otras. En este modelo se consideraron diferentes tipos de prominencia

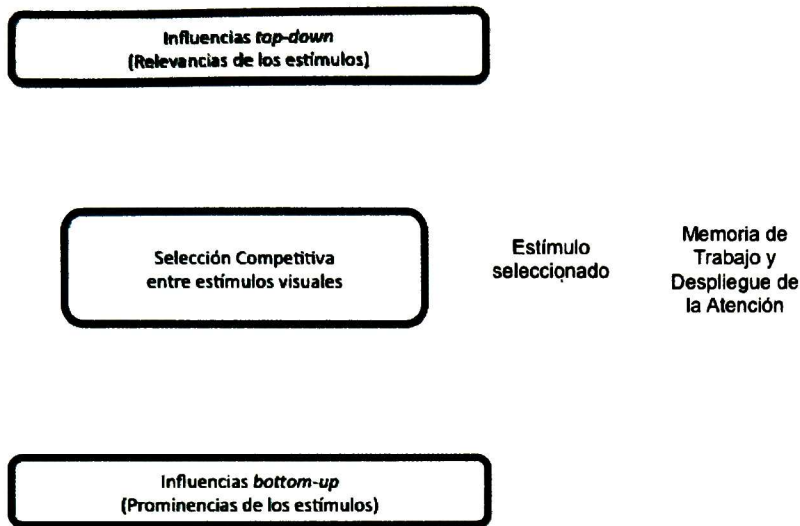


Figura 3.1: Diagrama general del modelo de atención propuesto en esta tesis.

acorde con el tipo de información que se procesa en los diversas áreas que comprenden la vía ventral. Así pues, se definió una prominencia basada en características físicas y otra basada en información semántica de los objetos en la escena. Más adelante se explica a detalle cómo se calculan y en base a qué información.

Con respecto a las influencias endógenas, se consideró sesgos en base a plantillas de atención (*attentional templates*) análogas a las de Desimone y Duncan (1995). Mediante estas plantillas la competencia entre los estímulos es sesgada hacia aquellos que se corresponden con dichas plantillas. En cierta medida los objetivos, metas e incluso algunas expectativas podrían expresarse en términos de plantillas. Estas plantillas incluyen información específica acerca de un estímulo en particular. Dicha información es proporcionada a los diferentes componentes de la vía ventral para que sesguen la competencia hacia estímulos cuyas propiedades se correspondan con las propiedades de la plantilla. A partir de esta información o sesgo *top-down* se calculan valores de relevancia para los estímulos. El término relevancia en este modelo indica el grado de importancia que tiene un estímulo visual acorde con el objetivo actual. De manera análoga a las prominencias, se consideran dos tipos principales de relevancias: relevancia basada en características físicas y relevancia semántica. Estas relevancias se calculan casi de manera simultánea que las prominencias correspondientes, es decir se calculan en base a la información de la imagen de entrada.

Al combinarse la información endógena (relevancia) y exógena (prominencia) de los estímulos se obtiene un valor de prioridad el cual es utilizado en nuestro modelo para determinar

el objeto que será atendido. El objeto con mayor prioridad será el objeto a atender. Nuestra aproximación considera dos formas de atender: atención abierta (*overt*) y cubierta (*covert*). En la atención abierta se realiza un movimiento de la mirada acorde con la ubicación del objeto con mayor prioridad y del centro de la mirada. La atención cubierta es un proceso mental que no requiere de realizar movimiento ocular alguno, simplemente se favorece el procesamiento del estímulo con mayor prioridad. Mientras el objeto con mayor prioridad es atendido, se le permite acceso a la memoria de trabajo donde otros procesos considerados de alto nivel utilizarán dicha información para su funcionamiento.

En nuestro modelo no consideramos el proceso de atención como un sistema centralizado sino como una propiedad emergente de un conjunto de procesos que se llevan a cabo en diferentes áreas y componentes del cerebro. Por lo tanto, para su diseño se identificaron diferentes áreas cerebrales y componentes involucrados en la atención así como el papel o función que realizan en dicho proceso. La figura 3.2 muestra el diagrama de conexiones de los diferentes componentes considerados para el modelo de atención que se propone. En la propuesta que presentamos, se hace especial énfasis en la función de la atención en la que actúa como un filtro de la enorme cantidad de información que constantemente los seres humanos recibimos de nuestros sentidos.

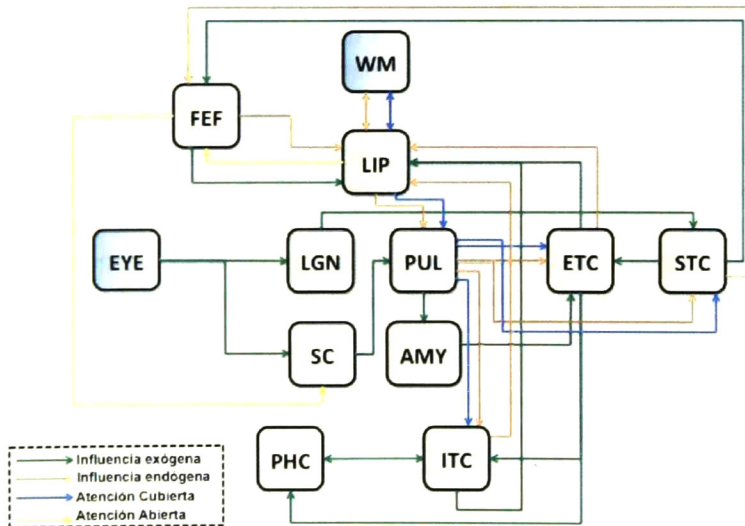


Figura 3.2: Componentes, áreas cerebrales y sus respectivas conexiones consideradas en la propuesta de modelo de atención visual.

### 3.3. Áreas Cerebrales y Componentes Involucrados

En la atención visual, diversas funciones y procesos están inmiscuidos en el desarrollo de la misma. En estos procesos involucrados tanto *bottom-up* como *top-down*, se obtienen diferentes datos e información acerca de los estímulos del ambiente que contribuyen a sesgar la atención en favor de cierto estímulo. Estas funciones en los seres humanos son llevadas a cabo por una serie de componentes y áreas cerebrales diferentes que mediante sus diversas conexiones e interacciones dan lugar a lo que conocemos como la atención visual.

Muchas áreas cerebrales han sido involucradas en el proceso de atención, las cuales se encuentran distribuidas tanto en la corteza cerebral como fuera de esta. Estos componentes no tienen una función única y exclusiva para el proceso de atención sino que también desempeñan otras labores para otros procesos cognitivos. Considerar todas las áreas cerebrales que se han identificado como activas durante experimentos de atención requeriría hacer un mapeo de una cantidad enorme de áreas del cerebro en nuestro modelo. En esta aproximación sólo se consideraron las áreas más importantes y representativas del proceso de atención visual; algunas de estas incluso representan un conjunto de estas o una abstracción de una función. Al considerar diversas áreas cerebrales, sus funciones y conexiones en este modelo de atención se pretende distribuir los diferentes procedimientos y procesos involucrados en la atención tal y como ocurre en las personas. También de esta forma se podrán añadir nuevas funcionalidades a dichos componentes conforme nueva evidencia vaya surgiendo o conforme nuevas interacciones e influencias de otros procesos cognitivos se añadan al modelo.

A continuación se enumera los diferentes componentes considerados para el modelo de atención que se propone en este trabajo. Además se hace una breve descripción de la función que desempeñan y de las conexiones consideradas en base a evidencia neurocientífica y de disciplinas afines. El nombre de los componentes se presenta en idioma inglés que es la forma más común en que estos son referidos en la mayor parte de la literatura.

#### EYE

El procesamiento visual comienza en la retina, la cual es un tejido sensible a la luz que recubre la superficie interna del ojo. La retina es una estructura en capas con varias capas de neuronas conectadas entre sí por las sinapsis. Las únicas neuronas que son directamente sensibles a la luz son las células fotorreceptoras. En la retina, las células fotorreceptoras hacen sinapsis directamente sobre las células bipolares, las que a su vez hacen lo mismo en las células ganglionares de la capa más externa. A través de sus células ganglionares y de sus interacciones en sus diversos circuitos, la retina envía flujos paralelos de información al cerebro, específicamente hacia el núcleo lateral geniculado y el colículo superior (Werblin y cols., 2001).

#### LGN

La mayor parte de las proyecciones retinales llegan al núcleo lateral geniculado o LGN. El LGN es un núcleo localizado en el tálamo y es considerado el centro de relevo primario de información visual. El LGN envía proyecciones a la corteza visual primaria para comenzar el procesamiento de la información visual en la corteza cerebral (Baluch y Itti, 2011).

### STC

La corteza visual primaria, corteza estriada o STC es la primer etapa cortical de procesamiento visual y se localiza en el polo posterior del lóbulo occipital. En la STC las neuronas con campos receptivos similares tienden a agruparse en columnas corticales (Wurtz y Kandel, 2000). Las neuronas de STC responden a atributos visuales simples como intensidad, color, orientación, entre otros (Li, 2002) y es una área modulada por la atención (Baluch y Itti, 2011). Acorde con la evidencia presentada en (Kastner y Ungerleider, 2000), las respuestas neuronales en STC son mejoradas para el estímulo que está siendo atendido, de tal forma que el procesamiento de información se sesga hacia este estímulo. La STC tiene proyecciones hacia la corteza extraestriada en la vía ventral, el colículo superior, los campos oculares frontales, entre otras áreas (Lund y Boothe, 1975).

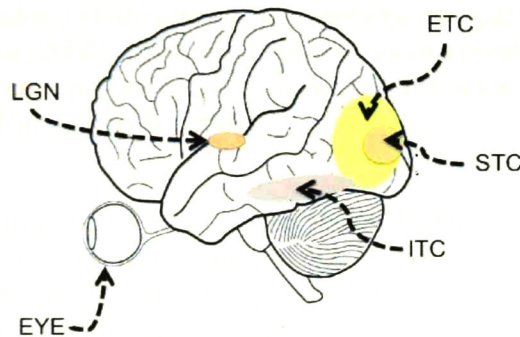


Figura 3.3: Áreas cerebrales y componentes involucrados en la atención visual (EYE, LGN, STC, ETC e ITC).

### ETC

La corteza extraestriada o ETC es la región de la corteza occipital localizada junto a la corteza visual primaria. Esta región comprende varias áreas funcionales entre las que se incluyen V3, V4, y V5. Las neuronas en ETC responden a la forma geométrica de los estímulos (VanRullen, 2003). Al estar localizada en un nivel más alto de la jerarquía de procesamiento de la vía ventral, el tamaño de sus campos receptivos se incrementa y son más sensibles a las características de los estímulos por lo que la modulación de la atención se incrementa. Al igual que la STC, la ETC favorece el procesamiento del estímulo atendido debido a la influencias *top-down* que recibe (Kastner y Ungerleider, 2000). Esta área tiene proyecciones



hacia la corteza inferotemporal, la corteza parahipocampal, el área lateral intraparietal, entre otras (Baluch y Itti, 2011), (Bar, 2004).

### **ITC**

La corteza inferotemporal o ITC se localiza en el lóbulo temporal y es considerada como el punto final de la jerarquía de procesamiento de la vía ventral visual. Se dice que la ITC funge como el almacenamiento central de objetos (Miyashita, 1993). Además se ha demostrado que esta corteza categoriza los objetos que se encuentran en la escena visual (Kiani y cols., 2007). De acuerdo con Kastner y Ungerleider (2000), en esta corteza la atención a un estímulo particular sesga la competencia por procesamiento posterior o por la entrada a memoria de trabajo. La ITC tiene proyecciones hacia la corteza parahipocampal, el área lateral intraparietal, entre otras áreas (VanRullen, 2003), (Fecteau y Munoz, 2006).

### **SC**

El colículo superior o SC es un área estructurada en capas localizada en el mesencéfalo. El SC recibe entradas de la STC y de la retina. Se dice que los eventos visualmente prominentes se representan en sus capas superficiales (Baluch y Itti, 2011). Se cree que en conjunto con los campos oculares frontales y el área lateral intraparietal comparten un mapa para guiar la atención (Fecteau y Munoz, 2006). El SC envía proyecciones al centro horizontal y vertical de la mirada para de esta forma proporcionar comandos apropiados para mover el (los) ojo(s) a una posición deseada (Purves y cols., 2001).

### **FEF**

Los campos oculares frontales o FEF es una región localizada en la corteza frontal que está fuertemente conectada con el área lateral intraparietal y el SC con las que se cree forma una red oculomotora. Se dice también que estas áreas mantienen un mapa que permite guiar la atención (Fecteau y Munoz, 2006). Las neuronas en la parte superior de esta región pueden controlar la producción de movimiento de los ojos a través de sus conexiones con el SC (Purves y cols., 2001). Acorde con lo expuesto en (Machado y Rafal, 2000), los FEF están involucrados en la determinación de la dirección y duración de la fijación de los ojos.

### **LIP**

El área lateral intraparietal o LIP (algunas veces llamada PEF (Shipp, 2004)) es una región localizada en la corteza parietal que contiene una representación de la prominencia relativa de los estímulos a lo largo de todo el campo visual. Esta área ha sido asociada al proceso de selección competitiva que determina que objeto será atendido (Knudsen, 2007), (Bisley y cols., 2011), (Baluch y Itti, 2011). El LIP mantiene fuertes conexiones con los FEF y el SC por lo que se cree comparte un mapa de atención y a la vez forman una red para generar movimientos de los ojos (Fecteau y Munoz, 2006), (Baluch y Itti, 2011). Se ha obtenido evidencia de que la actividad del LIP está estrechamente relacionada con la atención encubierta (*overt*). A través de sus proyecciones al pulvinar y de proyecciones directas, el

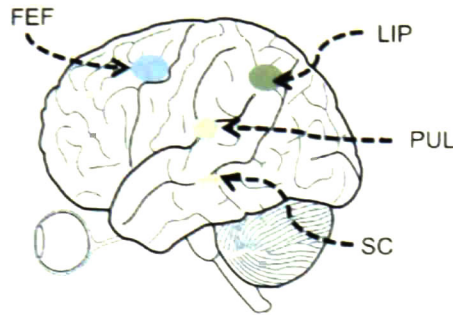


Figura 3.4: Áreas cerebrales involucradas en la atención visual (SC, FEF, LIP y PUL).

LIP envía sesgos *top-down* hacia las áreas de procesamiento en la vía ventral (Bisley y cols., 2011), (Kastner y Ungerleider, 2000).

### PUL

El pulvinar o PUL es un núcleo asociativo localizado en el tálamo, lo que indica que la mayor cantidad de información que recibe proviene de la corteza visual. El PUL ofrece una ruta para comunicación transcortical de manera indirecta. Esta área mantiene conexiones bidireccionales con muchas áreas entre las cuales se incluye la STC, ETC, ITC, LIP, SC, FEF, entre otras tantas (Shipp, 2004), (Baluch y Itti, 2011).

### PHC

La corteza parahipocampal o PHC es un conjunto de regiones localizadas en el lóbulo temporal medio. Se cree que la PHC mantiene una memoria de información relacionada con una tarea en específico. Este tipo de información podría ser la identidad o categoría del objetivo (VanRullen, 2003). A partir de la información proveniente de la ETC, la PHC podría hacer correspondencias entre representaciones visuales de baja frecuencia espacial (LSF, por sus siglas en inglés) y los marcos contextuales más probables. Estos marcos contienen información acerca de los posibles objetos que pueden encontrarse en ellos. La PHC mantiene fuertes conexiones con la ITC por lo que se cree que está involucrada también en el reconocimiento, familiaridad y categorización de los objetos (Bar, 2004).

### AMY

La amígdala o AMY es un conjunto de núcleos localizados dentro de los lóbulos temporales medios. La AMY ha sido identificada como una estructura cerebral clave que está involucrada en el procesamiento de estímulos emocionales. La AMY ha sido considerada como parte de un circuito rápido que registra el significado emocional de los estímulos en el ambiente. Evidencia indica que los estímulos visuales capturados por los humanos son proyectados desde la retina hasta la AMY a través de una vía tecto-pulvinar. Esta vía sub-cortical es rápida y

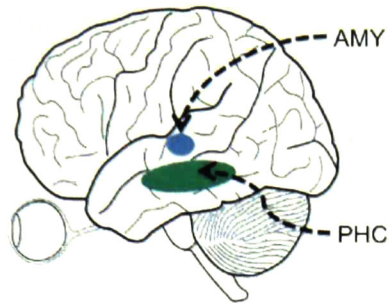


Figura 3.5: Áreas cerebrales involucradas en procesos que influyen en la atención (PHC y AMY).

transporta información visual burda lo que le permite a la AMY realizar evaluaciones rápidas pero imprecisas de los estímulos emocionalmente relevantes con el objetivo de preparar al organismo para contingencias emocionales (Vuilleumier, 2005). Acorde con la evidencia presentada en (Vuilleumier, 2005), áreas en la vía ventral mantienen interacciones recíprocas con la AMY, particularmente la ETC e ITC.

### WM

La memoria de trabajo o WM es un tipo de memoria de corto plazo. La WM mantiene una cantidad limitada de información por algunos segundos mientras esta información es evaluada. Se dice que los objetos compiten por tener acceso a la WM. La información que se mantiene en la WM sirve como base para tomar decisiones y la planeación de comportamientos complejos y además controla las señales *top-down* que modulan la sensibilidad de las representaciones neuronales que contribuyen a esta información. Evidencia sugiere que la memoria de trabajo es una función que se encuentra ampliamente distribuida en el cerebro (Knudsen, 2007), por tal motivo se decidió tener este componente y no todos las áreas cerebrales que podrían estar involucradas en él.

## 3.4. Influencias *Bottom-Up* en el Proceso de Atención

Los factores exógenos que influyen en el proceso de atención tienen que ver con las propiedades de los estímulos visuales. En nuestro modelo se propone cuantificar esta influencia *bottom-up* mediante una medida de prominencia para cada objeto en la escena visual. Considerando que las influencias exógenas en la atención se dan en base a diversas propiedades de los estímulos, se proponen distintas prominencias que consideren dichas propiedades. Estas prominencias se calculan en base a la información que es obtenida a lo largo de la jerarquía

de procesamiento de la vía ventral visual. Así pues, considerando las tres áreas principales de la vía ventral consideradas en nuestro modelo, se proponen dos prominencias principales. La primera prominencia se calcula considerando algunas características físicas de los estímulos. La segunda prominencia se calcula en base a información semántica de los estímulos y el contexto de la escena. En la figura 3.6 se muestra el diagrama general de las diferentes influencias *bottom-up* consideradas y los factores que intervienen en ellas.

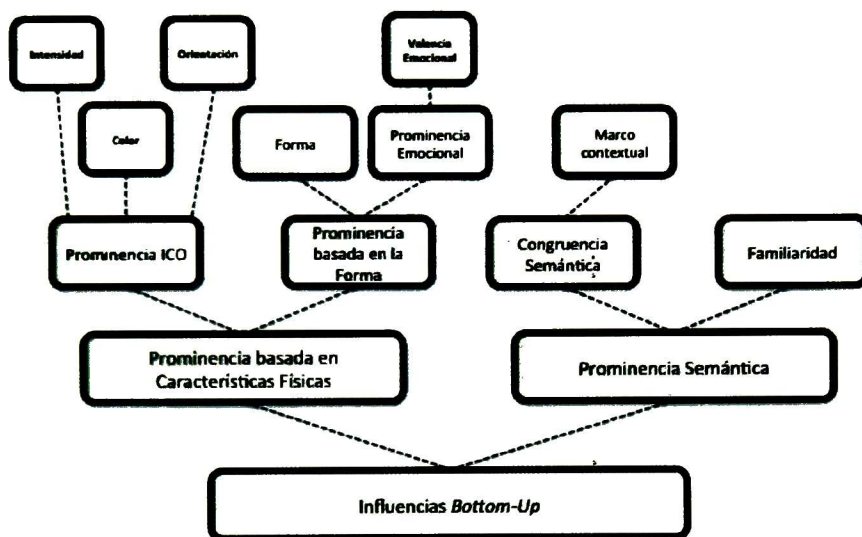


Figura 3.6: Influencias Exógenas (*bottom-up*) consideradas en la solución propuesta.

En esta sección se explica la forma de extraer las propiedades de los estímulos para calcular sus respectivas prominencias. Diferentes áreas cerebrales y componentes se encuentran involucrados en los diferentes mecanismos para la obtención de las prominencias de los estímulos. En este modelo cada área cerebral es representada como un componente con funcionalidad determinada en base a la evidencia descrita en la sección anterior. A través de figuras se muestran las diferentes conexiones y el flujo de información entre los diferentes componentes y por medio de tablas se describe la función de algunos de estos en las diferentes etapas del modelo.

### 3.4.1. Prominencia Basada en Características Físicas

La prominencia basada en características físicas que se propone en este modelo de atención considera información extraída de áreas en la vía ventral, la STC y ETC, e información proveniente de la AMY, en caso de que la escena contenga estímulos relevantes desde el punto

de vista emocional. El cálculo de esta prominencia se realiza en base a tres prominencias: prominencia basada en los atributos de intensidad, color y orientación (ICO), prominencia emocional y prominencia basada en la forma de los objetos.

En la figura 3.7 se presenta un diagrama con las diferentes áreas y componentes involucrados en el cálculo de estas tres prominencias y de la prominencia basada en características físicas. A continuación se presenta de manera detallada como se lleva a cabo el cálculo de esta prominencia y cómo se extraen los diferentes atributos y propiedades de la escena para calcular las tres prominencias atrás mencionadas. Cada una de estas tres prominencias representa una etapa en el modelo de atención visual.

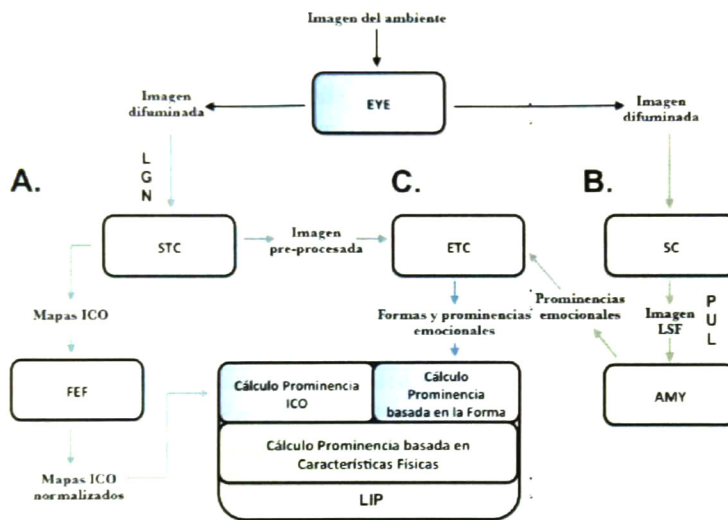


Figura 3.7: Áreas y componentes involucrados en la etapa de prominencia basada en características físicas. Prominencia ICO (A), Prominencia Emocional (B) y Prominencia basada en la forma (C).

### 3.4.1.1. Prominencia ICO

Esta prominencia se calcula en base a tres atributos simples: intensidad, color y orientación de cada pixel de la imagen de entrada. Durante esta etapa el componente EYE captura una imagen 2D que representa el ambiente de la criatura virtual. El componente EYE establece en esta imagen las áreas foveal y periférica. El área foveal se encuentra en el centro de la imagen y será de un tamaño pequeño predefinido. El área alrededor de esta es el área periférica la cual corresponderá al resto de la imagen. Enseguida se aplica una difuminación (*blur*) al área

periférica de tal forma que esta área tenga una menor resolución que el área foveal tal y como ocurre en la visión humana. Finalmente, el componente EYE envía esta imagen modificada hacia los componentes LGN y SC.

El componente LGN en este modelo se encarga simplemente de enrutar la información visual hacia el componente STC, que representa el primer componente que extrae características de la imagen. En el componente STC se calculan mapas ICO a partir de diferentes escalas de la imagen recibida del LGN. El mapa de intensidad tiene que ver con el contraste en la intensidad, el cual es detectado en los humanos por neuronas que son sensibles ya sea a centros oscuros con áreas circundantes brillantes o sensibles a la condición contraria (Leventhal, 1991). El mapa de color se obtiene utilizando un sistema de color de "doble-oponente" El centro de los campos receptivos de las neuronas es excitado por un color e inhibido por otro, mientras que lo contrario ocurre en el área circundante. Considerando esto, el componente STC crea mapas de forma simultánea para las oposiciones dobles de colores: rojo/verde, verde/rojo, azul/amarillo y amarillo/azul que existen en la corteza humana (Engel y cols., 1997). Con respecto al mapa de orientación, este se calcula utilizando filtros de Gabor en cuatro orientaciones diferentes. El componente STC calcula los mapas ICO de prominencia en base a mapas ICO en diferentes escalas espaciales. Una vez realizado esto, el componente STC envía estos mapas hacia el SC, y FEF. Cabe señalar que el componente STC realiza una especie de pre-procesamiento adicional a la imagen recibida. Dicho pre-procesamiento consiste en una clasificación de los píxeles de la imagen en clases de colores. Esta información se envía hacia la EST para de esta forma continuar con la jerarquía de procesamiento de la vía ventral. En la tabla 3.1 se muestra a detalle las funciones, entradas y salidas del componente STC en esta etapa del modelo.

Dado que no se ha encontrado evidencia de proyecciones desde la corteza estriada hacia el área lateral intraparietal, que es donde se dice existe una representación de la prominencia, el componente STC en nuestro modelo envía los mapas de prominencia ICO hacia el componente FEF. En nuestro modelo FEF normaliza estos mapas de prominencia para que los diferentes atributos puedan ser combinados de manera adecuada posteriormente. La tabla 3.2 muestra la función del componente FEF durante esta etapa del modelo.

En esta etapa el componente LIP calcula la prominencia ICO de cada píxel de la imagen capturada. Una vez que recibe los mapas de prominencia ICO normalizados, procede a calcular un único mapa de prominencia combinando la información de los mapas de intensidad, color y orientación. En la tabla 3.3 se muestra la función del componente LIP durante esta etapa.

Tabla 3.1: Función del componente STC en el cálculo de la prominencia ICO.

Componente	Entrada	Salida
LGN	Imagen 2D difuminada.	
ETC		Imagen 2D difuminada con clasificación de pixeles.
SC y FEF		Mapas de prominencia: $I, C, O$ .
Función		
<ol style="list-style-type: none"> <li>1. Clasifica cada pixel de la imagen de entrada en una de las <math>n_c</math> clases de colores.</li> <li>2. Crea escalas espaciales de la imagen desde la escala <math>\sigma \in [0 \dots 8]</math>.</li> <li>3. Establece las escalas del centro <math>c</math> y del área circundante <math>s</math>: <math>c \in \{2, 3, 4\}</math> y <math>s = c + \delta</math> con <math>\delta \in \{3, 4\}</math>.</li> <li>4. Calcula los mapas de características <math>I(c, s)</math>, <math>RG(c, s)</math>, <math>BY(c, s)</math> y <math>O(c, s, \theta)</math> con <math>\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}</math> por medio de operaciones de tipo centro-área circundante considerando las ecuaciones (1), (2), (3) y (4) propuestas en (Itti y cols., 1998), respectivamente.</li> <li>5. Calcula los mapas de prominencia de intensidad <math>I</math>, color <math>C</math> y orientación <math>O</math> como en las ecuaciones (5), (6) y (7) propuestas en (Itti y cols., 1998), respectivamente.</li> </ol>		

Tabla 3.2: Función del componente FEF's en el cálculo de la prominencia ICO.

Componente	Entrada	Salida
STC	Mapas $I, C, O$ .	
LIP		Mapas normalizados $N(I), N(C), N(O)$ .
Función		
<ol style="list-style-type: none"> <li>1. Para cada mapa (ya sea <math>I, C</math> o <math>O</math>) se realiza lo siguiente: <ol style="list-style-type: none"> <li>a) Normaliza los valores en el mapa en un rango fijo <math>[0 \dots M]</math>, donde <math>M</math> es el máximo global del mapa.</li> <li>b) Calcula el promedio <math>\bar{m}</math> de los máximos locales en el mapa.</li> <li>c) Multiplica el mapa por <math>(M - \bar{m})^2</math></li> </ol> </li> </ol>		

Tabla 3.3: Función del componente LIP en el cálculo de la prominencia ICO.

Componente	Entrada	Salida
FEF	Mapas normalizados $N(I), N(C), N(O)$ .	
<b>Función</b>		
1. Calcula la prominencia ICO: $S_{ICO} := \frac{N(I)+N(C)+N(O)}{3}$ , para cada pixel de la imagen.		

### 3.4.1.2. Prominencia Emocional

Experimentos han mostrado que los seres humanos tenemos una predisposición para atender elementos emocionalmente relevantes que se encuentran en el ambiente. Öhman y cols. (2001) sugiere que la habilidad para detectar de manera eficiente eventos significativos en el ambiente es esencial para propósitos de sobrevivencia. Acorde con sus estudios, ciertos tipos de estímulos tales como estímulos peligrosos y eventos emocionalmente relevantes pueden capturar de forma automática la atención de los individuos e interrumpir sus actividades cognitivas actuales. Considerando lo anterior, en este modelo se contempló una prominencia emocional de los estímulos en el ambiente que ayudará a sesgar la atención hacia ciertos estímulos que tengan una relevancia emocional considerable. A continuación se describe como es que se calcula esta prominencia emocional y las áreas involucradas en esta etapa. Cabe señalar que este proceso es parte del proceso cognitivo de emociones, es decir, en este modelo sólo se considera esta información (prominencia emocional) pero no los mecanismos ni la forma de evaluar los estímulos visuales para determinar cuando estos son emocionalmente significativos.

Anteriormente se mencionó que a través de una vía tecto-pulvinar se le proporciona a la AMY información para evaluar la significancia emocional de los estímulos de una forma rápida. No se conoce con precisión si el SC o el PUL o ambos modifican la información visual proveniente de la retina para que la AMY pueda realizar la acción mencionada. En nuestro modelo, se propone que el componente SC realiza una modificación de la imagen capturada para ayudar al componente AMY a evaluar los estímulos. De esta forma, el SC recibe la imagen del ambiente capturada por el componente EYE y le aplica una serie de operaciones para generar una imagen de frecuencia espacial baja o LSF. Las frecuencias espaciales bajas representan información global acerca de la forma y proporciones de los estímulos (Bar, 2004). Dicha imagen LSF es enviada hacia el componente PUL. El componente PUL, en esta etapa del modelo de atención que se propone y como parte de la vía tecto-pulvinar, actúa como un relevo de la información proveniente del componente SC. La imagen LSF recibida es reenviada hacia el componente AMY para continuar con el procesamiento de esta etapa.



En el componente AMY se realiza una evaluación de las formas de objetos extraídas de la imagen LSF que se recibe del PUL. Estas evaluaciones dan como resultado valencias emocionales que son utilizadas para calcular la prominencia emocional de los estímulos. En este modelo no se describe la manera en como se realiza esta evaluación ni como se obtienen las valencias emocionales para calcular la prominencia semántica, simplemente se considera dicha prominencia emocional en el proceso de atención. Los valores de prominencia emocional generados por la AMY son enviados hacia el componente ETC para que puedan influir en el proceso de atención. La tabla 3.4 esquematiza el componente AMY en esta etapa del modelo.

Tabla 3.4: Función del componente AMY en la prominencia emocional.

Componente	Entrada	Salida
PUL	Imagen LSF.	
ETC		Valores de prominencia emocional $S_{e_k}$ .
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Extrae las formas de los objetos <math>f_k</math> contenidos en la imagen LSF.</li> <li>2. Calcula la prominencia emocional <math>S_{e_k}</math> para cada objeto extraído <math>f_k</math> acorde con la información almacenada en términos de agradabilidad (<i>pleasantness</i>) y activación.</li> </ol>		

### 3.4.1.3. Prominencia Basada en la Forma

La prominencia basada en la forma se calcula considerando el tipo de información que se extrae en esta etapa de la jerarquía de procesamiento de la vía ventral. En esta etapa del modelo básicamente están involucrados dos componentes principales: ETC y LIP. A continuación se detalla la forma en que se extrae la información necesaria para calcular esta prominencia y la forma en que está se calcula.

El componente ETC se encarga de extraer la forma de los objetos contenidos en la imagen recibida del componente STC. Estas formas son extraídas mediante un algoritmo de segmentación que agrupa píxeles con propiedades similares en regiones. Enseguida estas formas se envían al componente ITC para continuar con el procesamiento en la vía ventral. En este modelo, ETC recibe valores de prominencia emocional provenientes del componente AMY y asocia tales valores con las formas correspondientes de los objetos extraídos. Posteriormente esta información es enviada hacia el componente LIP para que proceda a calcular la prominencia basada en la forma. En la tabla 3.5 se detalla la función del componente ETC en esta etapa del modelo.

Tabla 3.5: Función del componente ETC en el cálculo de la prominencia basada en la forma.

Componente	Entrada	Salida
STC	Imagen 2D difuminada con clasificación de píxeles.	
AMY	Valores de prominencia emocional $S_{e_k}$ .	
ITC y PHC		Formas de objetos $f_k$ .
LIP		Formas de los objetos $f_k$ y sus valores de prominencia emocional $S_{e_k}$ respectivos.
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Extrae la forma de los objetos <math>f_k</math> contenidos en la imagen difuminada mediante un algoritmo de segmentación.</li> <li>2. Realiza una correspondencia entre las formas extraídas <math>f_k</math> y los valores de prominencia emocional <math>S_{e_k}</math> con la finalidad de que cada forma extraída tenga asociado su valor emocional correspondiente.</li> </ol>		

A partir de la información enviada por el componente ETC, el componente LIP procede a calcular una prominencia basada en la forma de los objetos. Antes del cálculo de esta prominencia, el componente LIP asigna un valor de prominencia ICO a cada forma de objeto básica recibida. Esta prominencia ICO corresponde al promedio de las prominencias ICO de todos los píxeles que constituyen la forma. Enseguida, el componente LIP procede a calcular la prominencia basada en la forma considerando que tanto la forma de un objeto resalta con respecto a las formas de los objetos vecinos. Entre más difiera la forma de un objeto (física y/o emocionalmente) de la forma de sus objetos vecinos, mayor prominencia basada en la forma tendrá. La tabla 3.6 muestra la función del componente LIP durante esta etapa.

#### 3.4.1.4. Cálculo de la Prominencia Basada en Características Físicas

Una vez que se ha calculado la prominencia ICO y la prominencia basada en la forma (que lleva implícita la prominencia emocional) de los estímulos, se procede a calcular la prominencia basada en características físicas.

El componente LIP calcula la prominencia basada en características físicas básicamente como un promedio ponderado de la prominencia ICO y de la prominencia basada en la forma. El LIP asigna mayor peso a la prominencia (ICO o basada en la forma) cuando la diferencia entre esta y la media de las prominencias (ICO o basada en la forma) de los otros objetos es

Tabla 3.6: Función del componente LIP en el cálculo de la prominencia basada en la forma.

Componente	Entrada	Salida
ETC	Formas de los objetos $f_k$ y sus valores de prominencia emocional $S_{e_k}$ respectivos.	
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Calcula prominencia ICO para cada forma <math>f_k</math>: <math>S_{ICO_k} := \mu_{ICO_k}</math>, donde <math>\mu_{ICO_k}</math> es la media de las prominencias ICO de los pixeles contenidos en la forma <math>f_k</math>.</li> <li>2. Calcula la prominencia basada en la forma para cada objeto: <math>S_{S_k} := \Sigma D(f_k, V(f_k)) + \Sigma  S_{e_k} - V(S_{e_k}) </math>, donde la función <math>V</math> devuelve los objetos presentes en el vecindario de un objeto en particular, <math>D</math> es una función que devuelve un valor que indica que tan diferentes son dos formas de objetos dadas.</li> </ol>		

mayor que cierto umbral previamente establecido. En la tabla 3.7 se muestra de manera más precisa como el LIP calcula la prominencia basada en características físicas.

### 3.4.2. Prominencia Semántica

Acorde con la evidencia presentada en (Desimone y Duncan, 1995) el contexto temporal de los objetos en la escena visual influye en el proceso de atención ya que objetos que son nuevos o objetos conocidos pero que no han sido vistos recientemente suelen capturar la atención de manera automática. Otro factor que influye en la atención es la congruencia semántica de los objetos en la escena visual, es decir, si un determinado objeto es frecuentemente ubicado o asociado a un contexto particular. Se dice que objetos que son incongruentes con respecto al contexto de la escena tienden a ser atendidos (Underwood y Foulsham, 2006), (Võ y Henderson, 2009).

En la figura 3.8 se muestra un diagrama con las diferentes componentes involucrados en el cálculo de la prominencia semántica y los factores que intervienen en ella. Para influenciar la atención conforme a la evidencia, es necesario que los objetos sean reconocidos para así poder obtener información acerca de su congruencia semántica y familiaridad. Para esto se consideraron dos componentes principales: el ITC y PHC. El primero representa la última etapa de procesamiento de la vía ventral visual donde se dice se da el reconocimiento de los objetos. El segundo componente proporciona información acerca del contexto de la escena visual así como información acerca de la familiaridad de los objetos. El componente LIP se encarga de calcular la prominencia semántica de los objetos utilizando la información proporcionada por ambos componentes. Enseguida se explica con más profundidad la función

Tabla 3.7: LIP durante el cálculo de la prominencia basada en características físicas.

Función
<p>1. Interpola los valores de prominencia <math>S_{S_k}</math> en el rango <math>[min_{S_k}, max_{S_k}]</math> hacia valores en el rango <math>[min_{ICO_k}, max_{ICO_k}]</math>, donde <math>min_{S_k}</math>, <math>max_{S_k}</math>, <math>min_{ICO_k}</math>, <math>max_{ICO_k}</math> son los valores mínimo y máximo de prominencia basada en la forma y los valores mínimo y máximo de prominencia ICO de los objetos, respectivamente.</p> <p>2. Calcula <math>\mu_S</math>: la media de las prominencias <math>S_{S_k}</math> basadas en la forma de los objetos.</p> <p>3. Calcula <math>\mu_{ICO}</math>: la media de las prominencias ICO <math>S_{ICO_k}</math> de los objetos.</p> <p>4. Para cada objeto calcula la prominencia basada en características físicas <math>S_{F_k}</math> de la manera siguiente:</p> <p style="padding-left: 40px;">Si <math>S_{ICO_k} - \mu_{ICO} &gt; \delta_1</math> entonces <math>w_{ICO_k} := \frac{(S_{ICO_k} + \delta_1)}{S_{ICO_k}}</math>, si no <math>w_{ICO_k} := 1</math></p> <p style="padding-left: 40px;">Si <math>S_{S_k} - \mu_S &gt; \delta_2</math> entonces <math>w_{S_k} := \frac{(S_{S_k} + \delta_2)}{S_{S_k}}</math>, si no <math>w_{S_k} := 1</math></p> <p style="padding-left: 40px;"><math>S_{F_k} := \frac{(w_{ICO_k} S_{ICO_k} + w_{S_k} S_{S_k})}{2}</math>, donde <math>\delta_1</math> y <math>\delta_2</math> son umbrales predefinidos.</p>

exacta de cada componente en el modelo de atención propuesto.

### 3.4.2.1. Congruencia Semántica

El componente PHC determina el marco contextual correspondiente a la escena visual. Este contexto servirá para determinar la congruencia semántica de los objetos contenidos en la imagen. Note que la función para extraer el marco contextual de una escena visual tiene que ver con el proceso de memoria ya que la PHC junto con otras áreas como el hipocampo tienen información del marco(s) contextual(es) con los que un determinado objeto puede ser asociado. En este modelo no se detalla la forma en que está información está organizada ni como se extrae, simplemente se supone la existencia de una especie de base de datos con información de los objetos conocidos y el (los) marco(s) contextual(e) con el(los) que ha sido asociado. Una vez que el componente PHC obtiene el marco contextual de la escena, este envía al ITC dicho marco y un conjunto de objetos asociados al mismo. La tabla 3.8 muestra el funcionamiento del componente PHC durante esta etapa.

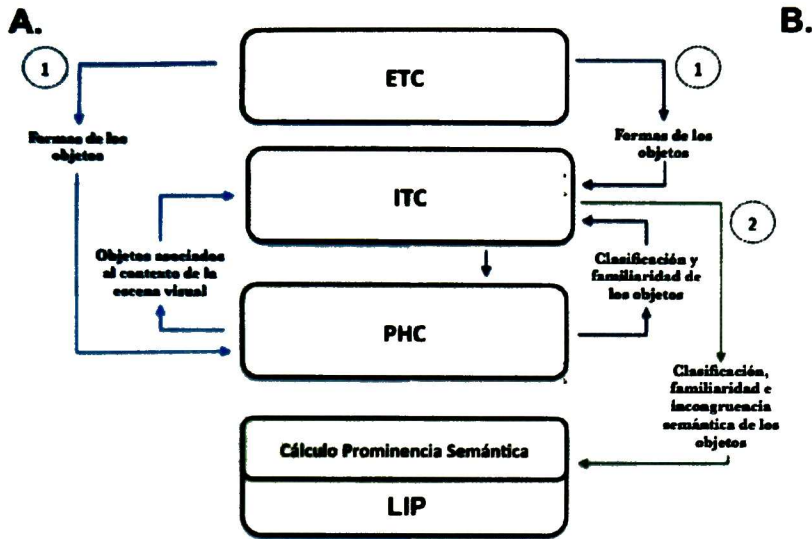


Figura 3.8: Áreas y componentes involucrados en la etapa de prominencia semántica. Congruencia Semántica (A), Familiaridad de los objetos (B).

Tabla 3.8: Función del componente PHC en la etapa de congruencia semántica.

Componente	Entrada	Salida
ETC	Formas de objetos $f_k$ .	
ITC		Conjunto de objetos asociados $P$ .
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Por cada forma de objeto <math>f_k</math> extrae de la base de datos el (los) marcos contextual(es) <math>M_k</math> asociados a él.</li> <li>2. Obtiene el marco contextual máximo <math>m_k</math> con el número máximo de objetos asociados a él.</li> <li>3. De la base de datos, obtiene a partir del marco contextual <math>m_k</math> el conjunto de objetos <math>P</math> asociados a dicho marco.</li> <li>4. En caso de que no exista un único marco contextual máximo, entonces <math>P := \emptyset</math>.</li> </ol>		

**3.4.2.2. Familiaridad de los Objetos**

El componente ITC en nuestro modelo envía al componente PHC las formas de los objetos contenidos en la imagen de entrada que fueron recibidos del componente EST. Esto con el fin de que se pueda obtener la clasificación de los objetos asociados a dichas formas. A partir de la información recibida de ITC, el componente PHC obtiene una clasificación de la formas de los objetos. Esta clasificación contendrá además un indicador acerca de la familiaridad de los objetos. En la tabla 3.9 se detalla este procedimiento. Cabe señalar que el proceso para obtener la clasificación de los objetos no se detalla en este modelo ya que se da por sentado que se cuenta con un mecanismo para llevar a cabo esta función.

Tabla 3.9: Función del componente PHC en la etapa de familiaridad de los objetos.

Componente	Entrada	Salida
ITC	Formas de objetos $f_k$ .	Conjunto de clasificaciones $O$ .
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Por cada forma de objeto <math>f_k</math> obtiene la clase de objeto <math>c_k</math> y su familiaridad correspondiente <math>h_k</math>, donde <math>h_k \in \{0, 1\}</math>. Cuando <math>h_k = 0</math> indica que el objeto no es familiar, mientras que cuando <math>h_k = 1</math> el objeto es familiar.</li> <li>2. Crea el conjunto <math>O</math> que contiene las clasificaciones de los objetos. <math>O := \{o_k \mid o_k = (c_k, h_k)\}</math>.</li> </ol>		

**3.4.2.3. Cálculo de la Prominencia Semántica**

La prominencia semántica se calcula en base a dos valores: la congruencia semántica y la familiaridad de los objetos extraídos de la imagen de entrada. El componente ITC en este modelo es el encargado de calcular la congruencia semántica de los objetos a través de la información recibida del componente PHC. Posteriormente, envía información de congruencia semántica y familiaridad de los objetos al componente LIP para que este proceda a calcular la prominencia semántica de los mismos. Enseguida se detalla la manera en que esta prominencia se calcula.

El componente ITC está a cargo de determinar la congruencia semántica de los objetos en la escena. Para llevar a cabo esto, el ITC verifica si los objetos identificados están asociados al marco contextual enviado por el componente PHC. La tabla 3.10 muestra la función del ITC en esta etapa del modelo. Una vez que se determina la congruencia (o incongruencia) de los objetos se envía esta información junto con la familiaridad de los mismos hacia el componente LIP para su procesamiento posterior.

Tabla 3.10: Función del componente ITC en el cálculo de la prominencia semántica.

Componente	Entrada	Salida
PHC	Conjuntos $P$ y $O$ .	
LIP		Conjunto $O'$ .
<b>Función</b>		
<p>1. Por cada clasificación <math>o_k</math> en <math>O</math>:</p> <ul style="list-style-type: none"> <li>▪ Si <math>o_k \in P</math> ó <math>P = \emptyset</math>, entonces <math>g_k := 0</math>, si no <math>g_k := 1</math>, donde <math>g_k</math> se define como un valor de incongruencia semántica.</li> <li>▪ Crea la tupla <math>o'_k</math> en base a la tupla <math>o_k</math> y el valor de incongruencia semántica:  <math>o'_k := (c_k, h_k, g_k)</math></li> </ul> <p>2. Crea el conjunto <math>O'</math> que contiene información semántica de los objetos: <math>O' := \{o'_k \mid o'_k = (c_k, h_k, g_k)\}</math>. Cada elemento de este conjunto contiene la clasificación (en base a la forma <math>f_k</math> del objeto), la familiaridad y un valor de incongruencia semántica de un objeto determinado.</p>		

El cálculo de la prominencia semántica se lleva cabo en el componente LIP. En la tabla 3.11 se detalla la forma en que esta prominencia se calcula. Si un objeto no es familiar para la criatura, mayor valor de prominencia semántica tendrá. De igual forma, tendrá mayor valor de prominencia semántica aquel objeto que no sea congruente con el contexto de la escena visual.

Tabla 3.11: Función del componente LIP en el cálculo de la prominencia semántica.

Componente	Entrada	Salida
ITC	Conjunto $O'$ .	
<b>Función</b>		
<p>1. Calcula la prominencia semántica <math>S_{M_k}</math> para cada elemento <math>o'_k \in O'</math>: <math>S_{M_k} := h_k + g_k</math>.</p>		

### 3.5. Influencias *Top-Down* en el Proceso de Atención

Dejando de lado por el momento las influencias exógenas, en esta sección se describen las influencias endógenas o *top-down* que fueron consideradas en este modelo de atención. Una de las influencias *top-down* más importantes que pueden influir en la atención de las personas

son los objetivos. Por ejemplo, en una tarea de búsqueda visual, el objetivo es encontrar en una imagen cierto objeto con características determinadas. Este objeto puede considerarse el objetivo en esta tarea, por lo tanto, la atención deberá sesgarse hacia estímulos u objetos que tengan características idénticas o similares a las del objetivo.

En este modelo de atención se consideró a los objetivos de la criatura como influencias endógenas. La manera que estas influencias sesgan el proceso de atención en el modelo se explica en la siguiente sección.

### 3.5.1. Plantillas de Atención

La mayor parte de los objetivos de una persona sólo pueden ser representados de manera abstracta. Por ejemplo, ganar una competencia, alimentarse para sobrevivir, escapar de una situación de peligro, entre otros. Estos objetivos de alto nivel son difíciles de expresar en términos tangibles. En este modelo no se considera este tipo de objetivos ya que representan objetivos complejos que muchas de las veces son compuestos de objetivos más sencillos. Objetivos sencillos tales como encontrar un determinado objeto en un experimento de búsqueda visual fueron considerados en este modelo. Si bien estos objetivos son sencillos tienen la ventaja de ser tangibles o de poder expresarse en términos menos abstractos y complejos. Además se cree que a partir de estos objetivos simples podría extenderse la capacidad del modelo para considerar objetivos de mayor nivel de complejidad. La manera en que se representa estos objetivos primarios es a través del uso de plantillas de atención (*attentional templates*).

Las plantillas atencionales que se utilizan en este modelo están inspiradas en aquellas a las que se hace referencia en (Desimone y Duncan, 1995). Acorde con lo expuesto en ese trabajo, las plantillas atencionales son un tipo de descripción, en memoria a corto plazo, que son utilizadas para sesgar la competencia en el sistema visual. Estas plantillas pueden ser vistas como un aspecto de memoria de trabajo y pueden especificar cualquier propiedad de un objeto tales como forma, color, ubicación, etc.

Considerando lo anterior, en este modelo se utilizan plantillas para representar los objetivos de una tarea en particular, como por ejemplo un objeto a buscar en un experimento de búsqueda visual. Estas plantillas son descripciones de características físicas tales como color, orientación forma e información semántica (tipo de objeto). La finalidad de tener plantillas conformadas con estas descripciones es la de poder simular influencias *top-down* en las áreas que forman parte de la vía ventral acorde con el tipo de información que extraen o a la cual responden. A partir de la información de estas plantillas se calculan valores de relevancia para los estímulos acorde con la correspondencia entre estos, sus atributos y las plantillas de atención. La relevancia de los estímulos resultaría en una competencia por atención en las áreas de la vía ventral visual que ya no estaría influenciada sólo por influencias exógenas sino también por influencias endógenas. En la figura 3.9 se muestra un diagrama con las diferentes



relevancias consideradas en este modelo de atención y los factores que se toman en cuenta para su cálculo.

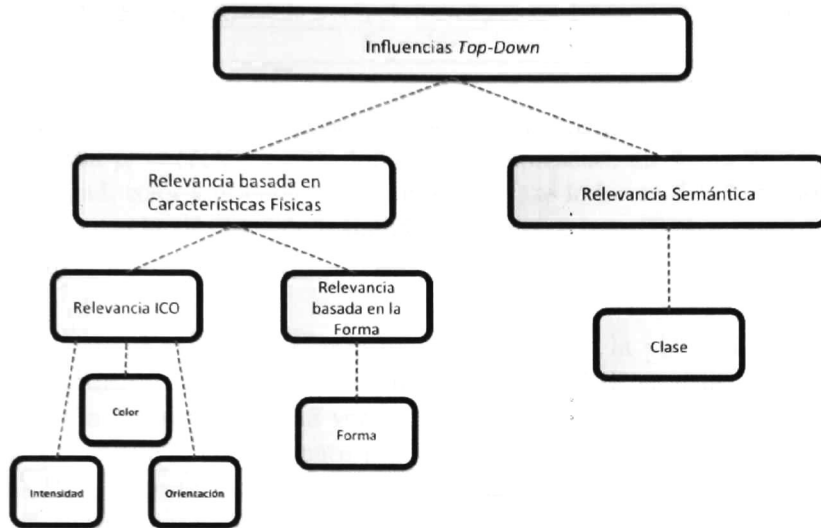


Figura 3.9: Influencias Endógenas (*top-down*) consideradas en la solución propuesta.

A continuación se describe el proceso de sesgo *top-down* en el modelo de atención que se propone, es decir, cómo se calcula la relevancia de los estímulos a partir de la información de las plantillas de atención y las propiedades de los mismos

Previamente se definió a la memoria de trabajo y la razón para considerarla como un componente y no como un conjunto de áreas cerebrales que realizan tal función. Además, también se mencionó que las plantillas atencionales podrían ser consideradas como un aspecto de memoria de trabajo. Por tal motivo en este modelo se decidió alojar estas plantillas atencionales en el componente WM para que a través de este, la información de estas plantillas o sesgo sea enviado hacia las áreas de procesamiento visual en la vía ventral. De esta manera, podemos decir que en este modelo la información alojada en el componente WM influye en el proceso de atención.

En el modelo de atención que se propone, el componente WM envía la plantilla de atención hacia el componente LIP para que se distribuya esta información en las áreas donde los estímulos compiten por procesamiento. La tabla 3.12 contiene información más precisa al respecto.

Considerando la evidencia que demuestra que el LIP está inmiscuido en el proceso de selección competitiva (Knudsen, 2007), es decir, en la selección del estímulo que entrará en la memoria de trabajo; en este modelo se propone que el componente LIP reciba sesgos *top-*

Tabla 3.12: Función del componente WM.

Componente	Entrada	Salida
LIP	Objeto que ganó la competencia $o_{winner}$ .	Plantilla $p_i$ .
<b>Función</b>		
<p>1. A partir de la información del objeto que ganó la competencia <math>o_{winner}</math> se construye una plantilla <math>p_i = (ICO_i, f_i, c_i)</math>, donde <math>ICO_i</math> representa un vector con los atributos de intensidad, color y orientación del objetivo, <math>f_i</math> es la forma del objeto asociado a la plantilla y <math>c_i</math> es la clasificación o tipo de objeto.</p>		

*down* del componente WM, es decir, información acerca de la plantilla del objeto. Esto se justifica desde el punto de vista de que la información en LIP es utilizada para determinar el ingreso a memoria de trabajo y una vez ahí se generan sesgos endógenos. En este modelo, la información que recibe el componente LIP de WM es dirigida hacia el componente PUL. El componente PUL en este modelo es utilizado para transmitir información transcortical. De tal forma, el componente PUL es utilizado en nuestro modelo como un enrutador de las plantillas atencionales provenientes del componente LIP que a su vez recibió del componente WM. La información de las plantillas que recibe PUL es distribuida entre las áreas de la vía ventral (tabla 3.13) con la finalidad de sesgar la competencia entre los estímulos hacia aquellos que cumplan o se correspondan con la información de la plantilla de atención.

Tabla 3.13: Función del componente PUL.

Componente	Entrada	Salida
LIP	Plantilla $p_i$ .	
STC		Vector $ICO_i$ .
ETC		Forma $f_i$ .
ITC		Clase $c_i$ .
<b>Función</b>		
<p>1. Distribuye entre las áreas de procesamiento en la vía ventral información de la plantilla de atención <math>p_i = (ICO_i, f_i, c_i)</math> acorde con el tipo de información que estas procesan.</p>		

### 3.5.2. Relevancia Basada en Características Físicas

La relevancia basada en características físicas que se sugiere en este modelo considera dos relevancias para su cálculo: relevancia ICO y relevancia basada en la forma. De manera

análoga a la prominencia basada en características físicas, esta relevancia refleja la importancia que tienen los estímulos visuales desde el punto de vista del objetivo que se persigue. Concretamente, esta relevancia es una medida de la importancia que tendrían los estímulos para el objetivo actual considerando sólo sus atributos físicos. Entre mayor coincidencia física tenga un estímulo con el objetivo actual, mayor será la relevancia basada en características físicas que tendrá.

La figura 3.10 muestra un diagrama con las diferentes áreas y componentes involucrados en el cálculo de la relevancia basada en características físicas. Enseguida se ahonda más en la forma en que esta relevancia se calcula, pero antes de eso, se detalla la forma en que las dos relevancias en las que se basa esta se calculan. En resumen, en las áreas de procesamiento visual, STC y ETC, se realiza una correspondencia entre las características físicas de los objetos contenidos en la imagen y las características físicas de la plantilla que representa el objetivo en cuestión. La información acerca de estas correspondencias se envía la componente LIP para que este se encargue de calcular la relevancia basada en atributos físicos de los diferentes objetos contenidos la imagen de entrada.

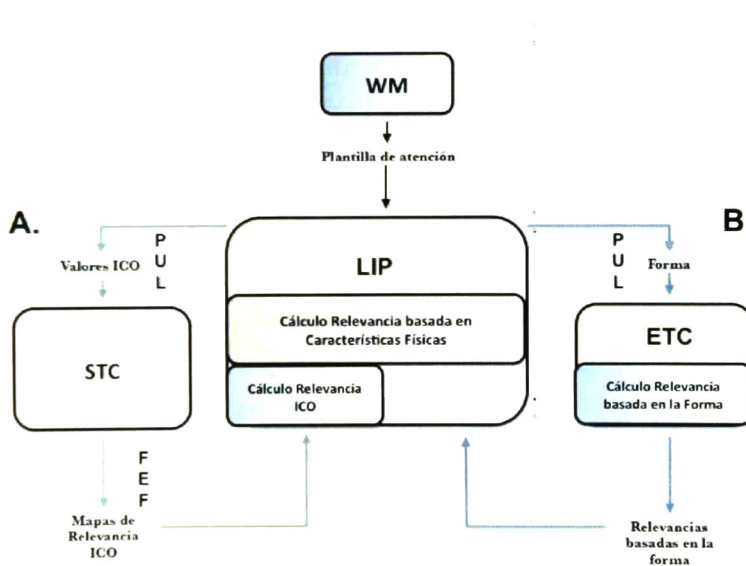


Figura 3.10: Áreas y componentes involucrados en la etapa de relevancia basada en características físicas. Relevancia ICO (A), Relevancia basada en la forma (B).

**3.5.2.1. Relevancia ICO**

La relevancia ICO, en particular, mide el grado o que tanto los atributos simples de un estímulo se asemejan a los atributos de un objeto que representa el objetivo de la tarea actual. En este apartado se explica la manera en que se calcula la relevancia de los estímulos considerando atributos simples tales como intensidad, color y orientación. El cálculo de esta relevancia involucra a dos componentes de este modelo, el STC y LIP, cuya función se describe enseguida.

El componente STC en este modelo recibe información de la plantilla correspondiente a los atributos físicos del objetivo: intensidad, color y orientación. Utilizando esta información, el componente STC realiza una correspondencia entre los atributos de la plantilla y los atributos de los estímulos que están o serán procesados en un muy corto plazo. Lo anterior con la finalidad de sesgar la competencia a favor del o los estímulos que tengan propiedades similares a las de la plantilla. Estas correspondencias se traducen en un mapa de relevancia para cada tipo de atributo (ICO). Los mapas de relevancia se envían al componente FEF para su posterior combinación. En la tabla 3.14 se muestra este procedimiento de manera más detallada.

**Tabla 3.14: Función del componente STC en el cálculo de la relevancia ICO.**

<b>Componente</b>	<b>Entrada</b>	<b>Salida</b>
PUL	Vector de atributos $ICO_i$ .	
FEF		Mapas de relevancia $R_I, R_C$ y $R_O$ .
<b>Función</b>		
<p>1. Calcula los mapas de relevancia de intensidad <math>R_I</math>, de color <math>R_C</math> y de orientación <math>R_O</math> en base a la información de los atributos simples de la plantilla <math>ICO_i</math> y a los mapas ICO generados a partir de la imagen de entrada:</p> <ul style="list-style-type: none"> <li>▪ Si <math>I(x, y) = ICO_i(I)</math>, entonces <math>R_I(x, y) := 1</math>, si no <math>R_I(x, y) := 0</math>.</li> <li>▪ Si <math>C(x, y) = ICO_i(C)</math>, entonces <math>R_C(x, y) := 1</math>, si no <math>R_C(x, y) := 0</math>.</li> <li>▪ Si <math>O(x, y) = ICO_i(O)</math>, entonces <math>R_O(x, y) := 1</math>, si no <math>R_O(x, y) := 0</math>.</li> </ul>		

El componente FEF en esta etapa recibe la información de los mapas de relevancia ICO que son enviados desde el componente STC y simplemente redirige estos mapas hacia el componente LIP. El componente LIP es el encargado de construir un único mapa de relevancia ICO a partir de los tres mapas enviados desde el componente STC por medio del componente FEF. La relevancia ICO es un promedio de las relevancias de intensidad, color y orientación (vease tabla 3.15).

Tabla 3.15: Función del componente LIP durante la etapa de relevancia ICO.

Componente	Entrada	Salida
FEF	Mapas de relevancia $R_I$ , $R_C$ y $R_O$	
<b>Función</b>		
1. Calcula la relevancia ICO: $R_{ICO} := \frac{R_I + R_C + R_O}{3}$ , para cada pixel de la imagen.		

### 3.5.2.2. Relevancia Basada en la Forma

Este tipo de relevancia se calcula en base a la forma de los objetos contenidos en la entrada visual y la forma especificada en la plantilla de atención. En otras palabras, se realiza una comparación entre las formas de los objetos y la contenida en la plantilla de atención. Los componentes involucrados en el cálculo de la relevancia basada en la forma son el ETC y LIP. A continuación se explica la manera en que se calcula esta relevancia o influencia *top-down* en este modelo.

El sesgo que recibe el componente ETC, proveniente del componente PUL, se utiliza para influir en la competencia entre los estímulos en esta etapa de la vía ventral. El componente ETC realiza una correspondencia entre la forma del objeto de la plantilla y la forma de los objetos que están siendo procesados o que lo serán próximamente. De esta manera aquellos estímulos cuya forma se corresponda con la de la plantilla tendrán una influencia mayor y será más probable que ganen la competencia por la atención. Al igual que el componente STC, este componente envía al componente LIP información acerca de las relevancias de los estímulos. La tabla 3.16 detalla lo descrito en este párrafo.

### 3.5.2.3. Cálculo de la Relevancia Basada en Características Físicas

En esta sección se describe la forma en que se determina la relevancia de los estímulos en base a sus atributos físicos (ICO y forma). Al igual que la prominencia basada en características físicas, esta relevancia se calcula considerando las dos relevancias simples descritas en los apartados anteriores.

El componente LIP, en base a las relevancias basadas en la forma que recibió del componente ETC y en base a las relevancias ICO que él calculó, procede a calcular una relevancia basada en características físicas para cada objeto presente en la escena visual. En la tabla 3.17 se describe de manera precisa la forma en que se calcula esta relevancia.

Tabla 3.16: Función del componente ETC en el cálculo de la relevancia basada en la forma.

Componente	Entrada	Salida
PUL	Forma de objeto $f_i$ .	
LIP		Conjunto de relevancias $R_S$ .
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Crea el conjunto de relevancias basadas en la forma <math>R_S</math> a partir de las relevancias <math>r_{S_k}</math> calculadas de la manera siguiente:</li> <li>2. <math>R_S := \emptyset</math></li> <li>3. Para cada forma de objeto <math>f_k</math> contenida en la imagen: <ul style="list-style-type: none"> <li>▪ Si <math>f_k \approx f_i</math>, entonces <math>r_{S_k} := 1</math>, si no <math>r_{S_k} := 0</math>.</li> <li>▪ Añade <math>r_{S_k}</math> al conjunto <math>R_S</math></li> </ul> </li> </ol>		

Tabla 3.17: LIP durante el cálculo de la relevancia basada características físicas.

<b>Función</b>
<ol style="list-style-type: none"> <li>1. Calcula relevancia ICO para cada forma <math>f_k</math>: <math>R_{ICO_k} := \mu_{R_{ico}}</math>, donde <math>\mu_{R_{ico}}</math> es la media de las relevancias ICO de los pixeles contenidos en la forma <math>f_k</math>.</li> <li>2. Calcula <math>\mu_{R_S}</math>: la media del conjunto <math>R_S</math> de prominencias basadas en la forma de los objetos.</li> <li>3. Calcula <math>\mu_{R_{ICO}}</math>: la media de todas las prominencias ICO <math>R_{ICO_k}</math> de los objetos.</li> <li>4. Para cada objeto calcula la prominencia basada en características físicas <math>R_{F_k}</math> de la manera siguiente: <math display="block">\text{Si } R_{ICO_k} - \mu_{R_{ICO}} &gt; \epsilon_1 \text{ entonces } z_{ICO_k} := \frac{(R_{ICO_k} + \epsilon_1)}{R_{ICO_k}}, \text{ si no } z_{ICO_k} := 1.</math> <math display="block">\text{Si } R_{S_k} - \mu_{R_S} &gt; \epsilon_2 \text{ entonces } z_{S_k} := \frac{(R_{S_k} + \epsilon_2)}{R_{S_k}}, \text{ si no } z_{S_k} := 1.</math> <math display="block">R_{F_k} := \frac{(z_{ICO_k} R_{ICO_k} + z_{S_k} R_{S_k})}{2}, \text{ donde } \epsilon_1 \text{ y } \epsilon_2 \text{ son umbrales predefinidos.}</math> </li> </ol>

### 3.5.3. Relevancia Semántica

El otro tipo de relevancia que se considera en este modelo de atención es la relevancia semántica. Dicha relevancia se calcula en base a la información semántica de la plantilla de atención. Un objeto tendrá relevancia semántica cuando pertenezca a la misma clase o al mismo tipo de objeto que el representado en la plantilla. Enseguida, se extiende la descripción de como se lleva a cabo esta acción. Cabe señalar que el componente que en este modelo está involucrado en el cálculo de la relevancia emocional es el ITC. Al componente LIP se le envía esta información para su posterior interpretación y uso. La figura 3.11 muestra un diagrama con las áreas y componentes involucrados en el cálculo de la relevancia semántica.

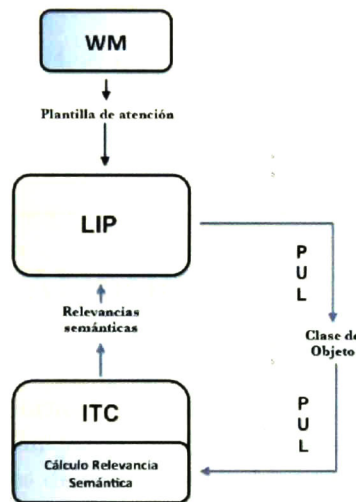


Figura 3.11: Áreas y componentes involucrados en la etapa de relevancia semántica.

#### 3.5.3.1. Cálculo de la Relevancia Semántica

La información que recibe el componente ITC por parte del PUL se utiliza para realizar una correspondencia entre los objetos, que están siendo procesados o que lo serán muy pronto, con la información de la plantilla de atención. El objeto cuya clasificación se corresponda con la de la plantilla tendrá más opciones de ser atendido que aquellos que no se corresponden. Es decir, dicho objeto tendrá mayor relevancia semántica que aquellos que no se corresponden con el tipo de objeto de la plantilla. Una vez realizado esto, se envía la información de estas relevancias al componente LIP (véase la tabla 3.18).

Tabla 3.18: Función del componente ITC en el cálculo de la relevancia semántica.

Componente	Entrada	Salida
PUL	Clase de objeto $c_i$ .	
LIP		Conjunto de relevancias $R_M$ .
<b>Función</b>		
<ol style="list-style-type: none"> <li>1. Crea el conjunto de relevancias basadas en la forma <math>R_M</math> a partir de las relevancias <math>r_{M_k}</math> calculadas de la manera siguiente:               <ol style="list-style-type: none"> <li>2. <math>R_M := \emptyset</math>.</li> <li>3. Para cada objeto <math>o_k</math> presente en la imagen:                   <ul style="list-style-type: none"> <li>▪ Si <math>c_k = c_i</math>, entonces <math>r_{M_k} := 1</math>, si no <math>r_{M_k} := 0</math>. Donde <math>c_k</math> es la clase asociada al objeto <math>o_k</math>.</li> <li>▪ Añade <math>r_{M_k}</math> al conjunto <math>R_M</math>.</li> </ul> </li> </ol> </li> </ol>		

## 3.6. Despliegue de la Atención

En esta última sección del capítulo se detalla la manera en que este modelo realiza el despliegue de la atención, es decir, la forma en que un objeto es atendido. Pero antes de esto, se explica como las influencias *bottom-up* y *top-down* se combinan para sesgar la competencia por atención. Una vez que ambos tipos de sesgos se combinan se determina cual es el objeto con mayor prioridad y se procede a atender tal objeto ya sea de manera abierta o cubierta. En la figura 3.12 se muestra un diagrama que contiene las diferentes áreas y componentes involucrados en el despliegue de la atención.

### 3.6.1. Cálculo de Prioridad y Selección Competitiva

El cálculo de la prioridad implica la integración de las influencias *bottom-up* y *top-down* explicadas en los apartados anteriores. La prioridad de un objeto en este modelo se hace mediante la combinación de información acerca de su prominencia y de su relevancia. En base a esta prioridad, se determina el objeto que será sujeto de atención.

Una vez que se han calculado la prioridad de los objetos en la escena visual, entonces se procede a seleccionar el objeto que será atendido. De manera lógica podría pensarse que siempre el objeto con mayor prioridad tendría que ser atendido y por lo tanto la selección del objeto a atender está en función siempre de la prioridad. Sin embargo, en este modelo de



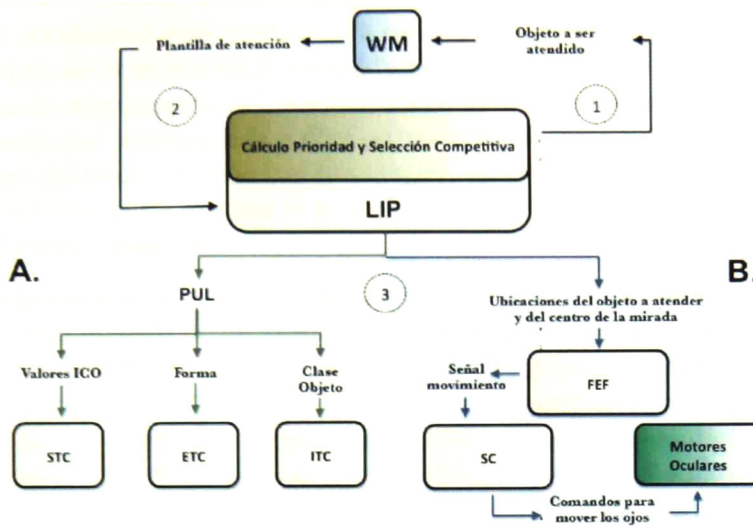


Figura 3.12: Áreas y componentes involucrados en el despliegue de la atención. (A) Atención cubierta (*Covert Attention*), (B) Atención abierta (*Overt Attention*).

atención se consideraron dos aspectos importantes.

En primer lugar se consideró la posibilidad de atender un objeto sin tener que calcular todo el mapa de prioridad lo cual requiere tiempo. En este modelo se decidió que cualquier objeto podría ser atendido si su valor de prominencia basada en características o de prominencia semántica es igual o mayor que un umbral predefinido. Durante el cálculo de la prominencia basada en características físicas, se ponderan las dos prominencias, es decir, se le da un peso mayor a cualquiera de estas cuando el valor de estas prominencias (ICO y basada en la forma) difieren del resto de manera considerable. Esta ponderación permite que un objeto pueda tener una prominencia basada en características bastante grande de tal forma que tal objeto pueda ser atendido de manera rápida sin la necesidad de considerar otros tipos de prominencia. En el caso de la prominencia semántica cuando un objeto no es familiar y no es congruente con el contexto de la escena, es decir, es un objeto nuevo, se tendrá la posibilidad de atender este objeto sin la necesidad de considerar el mapa de relevancia. El motivo de esta variación es que experimentos han señalado que objetos que son demasiado prominentes (físicamente o emocionalmente) generalmente capturan la atención incluso antes de ser reconocidos completamente. También, objetos que son desconocidos por las personas suelen ser atendidos de manera automática (Desimone y Duncan, 1995). Así pues, esta variación permite a nuestro modelo el poder exhibir dichos comportamientos. De manera análoga se considero un mecanismo similar para la relevancia basada en características físicas y la relevancia semántica. Es decir, siempre que cierto estímulo tenga una relevancia (basada en

características o semántica) muy contrastada con las del resto de los objetos, entonces dicho estímulo será sujeto de atención sin la necesidad de tener que calcular la prioridad de los objetos. Considerando este aspecto del modelo, surge un caso interesante: cuando un estímulo tiene una relevancia muy contrastada y otro estímulo tiene una prominencia muy contrastada con respecto a las del resto. Acorde con lo anterior, en este caso ambos objetos deberían ser atendidos. Para decidir la competencia en este caso se plantea utilizar una función *random* para decidir que objeto atender.

El segundo aspecto que se consideró en este modelo es el de añadir la posibilidad, si así se desea, de agregar una manera no determinista de elegir que objeto atender en base al mapa de prioridad. Muchos modelos de atención (por ejemplo (Itti y cols., 1998)), atienden los objetos de una imagen o escena visual en orden descendente de prominencia o de prioridad cuando incluyen influencias *top-down*. Esto resulta en un comportamiento idéntico cuando se trabaja con la misma imagen. Con el fin de tener la posibilidad de exhibir distintos comportamientos con una misma imagen, se determinó utilizar un mecanismo que permita una selección no determinista del objeto a atender. Este mecanismo lo que hace es que mediante una función *random* se determina el objeto a atender. Pero esta función asigna mayor probabilidad al objeto con mayor prioridad, mientras que a los objetos con menos prioridad les asigna un valor de probabilidad menor de ser seleccionados, es decir, se aplicó una especie de ruleta proporcional a la prioridad de los estímulos. Mediante esto, se tiene una forma no determinista de atender una imagen lo cual enriquece un poco más el modelo. Cabe señalar que el aspecto descrito en el párrafo anterior siempre se considera, mientras que este aspecto se considera sólo cuando el proceso llega hasta la etapa del cálculo del mapa de prioridad.

El componente a cargo de la integración de las influencias exógenas y endógenas es el componente LIP. Así pues, a partir de la información proveniente de las diferentes áreas de procesamiento en la vía ventral se calculan valores de prioridad para cada estímulo. Estos valores de prioridad se calculan en base a la integración de información de prominencia y relevancia, es decir, *bottom-up* y *top-down*, respectivamente. En nuestro modelo el componente LIP también es el encargado de hacer la selección del objeto a atender teniendo en cuenta los aspectos atrás mencionados y no sólo la prioridad de los objetos. En la tabla 3.19 se detalla las acciones del componente LIP durante el proceso de cálculo de prioridad y selección competitiva. Acorde con esta tabla, la selección competitiva se puede realizar en tres tiempos distintos: 1) al momento de tener valores de prominencia y relevancia basados en características, 2) al momento de tener información acerca de la prominencia y relevancia semántica y 3) después de calcular la prioridad de los objetos. Los pasos 1 y 2 de la tabla se llevan a cabo una vez que se calculan las prominencias y relevancias basadas en características y cuando se calculan las prominencias y relevancias semánticas, respectivamente. El paso 3 se realiza cuando se tiene información de prominencias y relevancias para calcular las prioridades de los objetos. El paso 4 se realiza inmediatamente después del paso 3.

Tabla 3.19: LIP durante el cálculo de prioridad y selección competitiva.

Componente	Entrada	Salida
WM		Objeto por atender $o_{winner}$
<b>Función</b>		
<p>1. Verifica si un objeto debe ser atendido en base a su prominencia o relevancia basada en características:</p> <ul style="list-style-type: none"> <li>▪ <math>o_{winner} := NULL</math>.</li> <li>▪ Si <math>max(S_{F_k}) \geq \gamma_{S_F}</math>, entonces <math>o_{w_S} := f_k</math>, donde <math>\gamma_{S_F}</math> es un umbral predefinido.</li> <li>▪ Si <math>max(R_{F_k}) \geq \gamma_{R_F}</math>, entonces <math>o_{w_R} := f_k</math>, donde <math>\gamma_{R_F}</math> es un umbral predefinido.</li> <li>▪ Si <math>o_{w_S} \neq o_{w_R}</math>, entonces <math>o_{winner} := random(o_{w_S}, o_{w_R})</math>. Si no <math>o_{winner} := o_{w_S}</math>.</li> </ul> <p>2. Verifica si un objeto debe ser atendido en base a su prominencia o relevancia semántica:</p> <ul style="list-style-type: none"> <li>▪ <math>o_{winner} := NULL</math>.</li> <li>▪ Si <math>max(S_{M_k}) \geq \gamma_{S_M}</math>, entonces <math>o_{w_S} := f_k</math>, donde <math>\gamma_{S_M}</math> es un umbral predefinido.</li> <li>▪ Si <math>max(R_{M_k}) \geq \gamma_{R_M}</math>, entonces <math>o_{w_R} := f_k</math>, donde <math>\gamma_{R_M}</math> es un umbral predefinido.</li> <li>▪ Si <math>o_{w_S} \neq o_{w_R}</math>, entonces <math>o_{winner} := random(o_{w_S}, o_{w_R})</math>. Si no <math>o_{winner} := o_{w_S}</math>.</li> </ul> <p>3. Para cada objeto <math>k</math> calcula su prioridad <math>P_k</math> de la manera siguiente:</p> <ul style="list-style-type: none"> <li>▪ <math>P_k := \frac{S_{F_k} + S_{M_k} + R_{F_k} + R_{M_k}}{4}</math>. Cabe señalar que previamente los valores de prominencia y relevancia debieron de haberse interpolado a un rango de valores común con el fin de que cada uno de estos valores contribuya de forma apropiada al cálculo de la prioridad.</li> </ul> <p>4. Selecciona el objeto a atender en base a su prioridad:</p> <p><math>o_{winner} := roulette(P)</math>, donde <math>roulette</math> es una función que selecciona el objeto a atender en base a una ruleta proporcional al conjunto de prioridades <math>P</math> de los objetos. Entre más grande sea el valor de prioridad <math>P_k</math> de un objeto mayor probabilidad tiene de ser elegido para ser atendido.</p>		

### 3.6.2. *Covert Attention*

La atención cubierta (*covert attention*) consiste en atender un estímulo sin la necesidad de dirigir ningún órgano sensorial hacia tal estímulo. Este tipo de atención consiste más en un enfoque mental (Baluch y Itti, 2011). En el caso específico de la atención visual, la atención encubierta consiste en atender un estímulo visual sin la necesidad de mover los ojos.

Con el fin de atender al objeto que ganó la competencia, se envían sesgos de atención hacia las áreas de procesamiento en la vía ventral para favorecer el procesamiento de dicho objeto, es decir, sesgar la competencia por atención. Los sesgos de atención pueden ser considerados como influencias *top-down* que sesgan la atención hacia cierto objeto descrito por una plantilla. Considerando esto, el objeto que gana la competencia puede ser considerado el objetivo actual del sistema ya que se desea que este sea atendido. De esta forma, podemos considerar el mismo proceso que el descrito en la sección Plantillas de este apartado para explicar la atención cubierta. Es decir, una vez que se hace la selección del objeto a atender en la escena visual, el componente LIP envía tal información al componente WM para que se genere una plantilla de atención para atender al objeto que ganó la competencia. Enseguida, el componente WM envía la plantilla atencional a LIP para que este sea el encargado de enviar los sesgos atencionales a las áreas de procesamiento visual de la vía ventral por medio del componente PUL. Por medio de estos sesgos de atención se favorecerá al objeto que ganó la competencia.

Dado que en este modelo se consideran imágenes estáticas, los sesgos de atención no se ven reflejados de manera explícita puesto que se trabaja con una imagen que no cambia ya que no aparecen nuevos objetos en ella que pudieran competir con los ya presentes. El caso donde sí se vería reflejado el sesgo es aquel donde se atiende a un objeto antes de calcular su prioridad. En este modelo, siempre el objeto que gana la competencia por atención se atiende de manera cubierta. Es decir, se favorece su procesamiento mediante sesgos de atención representados por plantillas.

### 3.6.3. *Overt Attention*

La atención abierta (*overt attention*) consiste en orientar los órganos sensoriales hacia estímulos de interés (Baluch y Itti, 2011). En el caso de atención visual, la atención abierta consiste en mover los ojos hacia cierta ubicación de interés. En este modelo, esta ubicación de interés está determinada por el objeto que ganó la competencia por atención. Cuando el objeto que fue seleccionado para ser atendido se encuentra fuera del área foveal de la imagen, entonces dicho objeto es atendido abiertamente, es decir, se realizará un movimiento de los ojos. Sin embargo, no sólo debe moverse los ojos hacia el objeto de interés sino que también se debe favorecer su procesamiento. En este modelo dicho favorecimiento se da a través de la atención cubierta.

Aunque el movimiento de los ojos debiera generar una nueva imagen del ambiente, en este modelo no se considera esta opción por ahora. El movimiento de los ojos, simplemente lo vemos como un comportamiento y no como un generador de una nueva entrada al sistema. Posteriormente, deberá ser considerada esta capacidad para tener un modelo de atención más preciso. A continuación se realiza una breve descripción de los componentes involucrados en el movimiento de los ojos.

Considerando la evidencia biológica, el componente FEF en nuestro modelo se encarga de evaluar si se requiere un movimiento de ojos para atender al objeto seleccionado. En la tabla 3.20 se muestra la función del FEF en esta etapa del modelo.

Tabla 3.20: Función del componente FEF en la atención abierta.

Componente	Entrada	Salida
LIP	Ubicaciones $l_{winner}$ y $l_{gaze}$ .	
SC		Señal de movimiento $s(l_{winner}, l_{gaze})$ .
Función		
<p>1. A partir de la ubicación del centro del objeto a atender <math>l_{winner}</math> y de la ubicación del centro de la mirada (centro del área foveal) <math>l_{gaze}</math>, se determina si se requiere un movimiento de ojos de la siguiente manera:</p> <ul style="list-style-type: none"> <li>▪ Si <math>l_{winner}[x]</math> o <math>l_{winner}[y]</math> está fuera del área foveal, entonces envía al componente SC la señal <math>s(l_{winner}, l_{gaze})</math>.</li> </ul>		

El componente SC se encarga de generar los comandos apropiados para el movimiento de los ojos si es que recibe una señal del componente FEF para realizar tal movimiento. Una vez que se generan estos comandos, el componente SC los envía a los motores de los ojos para su ejecución. La tabla 3.21 explica la forma en que se generan estos comandos.

### 3.6.4. IOR (Inhibition of Return)

Una vez que un objeto es atendido durante un tiempo fijo determinado, se procede a atender otro objeto en la escena visual. Con el fin de evitar que el mismo objeto sea atendido, diferentes modelos de atención (Itti y Koch, 2000), (Fecteau y Munoz, 2006) y otros más han considerado un mecanismo IOR que evita que un objeto previamente atendido vuelva a competir por atención. En este modelo se consideró un mecanismo IOR. El componente LIP mantiene una especie de lista con los objetos previamente atendidos, entonces cada vez que atiende un objeto lo agrega a la lista para que no vuelva a competir por atención. Los objetos

Tabla 3.21: Función del componente SC en la atención abierta.

Componente	Entrada	Salida
FEF	Señal de movimiento $s(l_{winner}, l_{gaze})$ .	
Motores de los ojos		Comandos $c_h(a_h, dir_h)$ y $c_v(a_v, dir_v)$ .
Función		
<p>1. Se crean comandos para mover los ojos de la siguiente manera:</p> <ul style="list-style-type: none"> <li>▪ <math>dist_h := l_{gaze}[x] - l_{winner}[x]</math>.</li> <li>▪ <math>dist_v := l_{gaze}[y] - l_{winner}[y]</math>.</li> <li>▪ Si <math>dist_h &gt; 0</math>, entonces <math>dir_h := right</math>, si no <math>dir_h := left</math>.</li> <li>▪ Si <math>dist_v &gt; 0</math>, entonces <math>dir_v := up</math>, si no <math>dir_v := down</math>.</li> <li>▪ <math>a_h :=  dist_h </math>, <math>a_v :=  dist_v </math>.</li> <li>▪ Crea comandos <math>c_h(a_h, dir_h)</math> y <math>c_v(a_v, dir_v)</math>.</li> </ul>		

que se encuentren en la lista de objetos atendidos no serán considerados para competir por la atención y el ingreso a memoria de trabajo.

# Capítulo 4

## Implementación y Resultados

### 4.1. Introducción

En el capítulo anterior se detalló la propuesta que se propone en esta tesis para modelar el proceso cognitivo de la atención. En esta sección se describen aspectos acerca de la implementación del modelo de atención visual y los casos de estudio propuestos para probar el funcionamiento de dicho modelo.

Sin duda la implementación de cualquier tipo de modelo computacional es importante así como los resultados arrojados por este cuando se prueba su funcionamiento en experimentos o casos de estudio específicos. Este capítulo de la tesis describe estos aspectos del modelo del sistema cognitivo de atención propuesto. Existe una diversidad de casos de estudio tanto psicológicos como neurocientíficos que son utilizados para probar la atención en seres humanos y animales, especialmente primates. En este trabajo se consideraron dos casos de estudio generales: *Free-Viewing* y *Visual Search*. Ambos experimentos han sido utilizados para probar el funcionamiento de modelos de atención en diversos trabajos (Itti y cols., 1998), (Itti y Koch, 2000), (Sun y cols., 2008), (Li, 2002), (Bisley y cols., 2011), entre muchos otros. A través de estos experimentos, se desea probar que el modelo de atención que se propone es capaz de lidiar con situaciones típicas y relativamente sencillas para que a partir de ahí se pueda extender y utilizar en experimentos de atención mucho más complejos y desafiantes.

El capítulo se encuentra organizado en cuatro secciones principales. La primera sección describe las herramientas computacionales utilizadas para la implementación del modelo de atención que se propone. La segunda y tercera sección plantean el primer y segundo caso de estudio, respectivamente, que se utilizarán para probar el funcionamiento y desempeño del modelo. Finalmente, en la última sección se realizan conclusiones a partir de los resultados obtenidos en los casos de estudio propuestos para este trabajo.

## 4.2. Herramientas Utilizadas

En esta sección se describen las herramientas de software que ayudaron en la implementación del modelo de atención propuesto y se especifica en que funciones o subprocesos fueron empleadas.

Al leer la descripción del modelo en el capítulo anterior cualquiera puede darse cuenta que en el modelo de atención intervienen diferentes subprocesos o subfunciones llevadas a cabo por diferentes componentes. En este trabajo no se intenta proponer nuevas formas para llevar a cabo todos y cada uno de los diferentes subprocesos involucrados en el proceso de atención ya que esto requeriría otras tesis y sería un trabajo demasiado grande y ambicioso. Con el fin de acotar el trabajo y reducir la complejidad del mismo, se optó por utilizar algunos mecanismos y funciones ya implementadas para recrear la función de ciertos subprocesos involucrados en la aproximación que se propone.

Cabe resaltar que aunque algunas herramientas ajenas a este trabajo son utilizadas en la implementación del modelo, esto no implica que nada del modelo propuesto en esta tesis es novedoso o que no se desarrolló software propio para la implementación del sistema de atención.

### 4.2.1. iNVT

iNVT es un conjunto de clases C++ para el desarrollo de modelos de visión neuromórficos. Es un proyecto a cargo del laboratorio *iLab* en la Universidad del Sur de California.

El proyecto iNVT tiene como objetivo desarrollar la próxima generación de algoritmos de visión. para ello. proporciona una base de software que está específicamente orientada hacia el desarrollo de modelos y sistemas neuromórficos.

Algunas características de este conjunto de bibliotecas de software son las siguientes:

- Clases auxiliares de bajo nivel entre las cuales se incluyen clases para representar puntos 2D, pixeles RGB, rangos, rectángulos, temporizadores, entre otras.
- Plantillas para imágenes y conjuntos de imágenes con diversas funciones para su procesamiento.
- Funciones para el manejo de imágenes tales como lectura/escritura en varios formatos.
- Clases neuromórficas entre las que se incluye una jerarquía de canales de atributos de bajo nivel para el cálculo de mapas neuronales que responden a intensidad, color y orientación, etc.



Aunque iNVT tiene un conjunto de funcionalidades para recrear diferentes aspectos de la atención, sólo algunas funciones y clases de este software fueron utilizadas para crear los mapas de prominencia de intensidad, color y orientación, que en el modelo propuesto, son generados por el componente STC. Como se mencionó en la tabla 3.1, este componente calcula los mapas de prominencia en base a lo propuesto en el trabajo Itti y cols. (1998), cuya forma de calcularse se encuentra implementada en iNVT.

El código de la clases es distribuido de manera libre por lo que puede ser utilizado para fines académicos sin problema alguno.

#### 4.2.2. OpenCV

OpenCV (*Open Computer Vision*) es una biblioteca libre de visión artificial que fue inicialmente desarrollada por Intel. Esta biblioteca incluye una gran cantidad de algoritmos de visión por computadora que han sido utilizados en diversas aplicaciones tales como sistemas de seguridad para detectar movimientos, reconocimiento de objetos, entre otras.

OpenCV cuenta con una estructura modular lo cual significa que está compuesta por varias bibliotecas estáticas y compartidas. Algunos de los módulos de OpenCV son:

- core: el cual contiene definiciones de estructuras de datos básicas para el procesamiento de imágenes.
- imgproc: es un módulo que incluye funciones para el filtrado lineal y no lineal de imágenes, transformaciones geométricas, conversión entre diferentes espacios de colores, histogramas, entre otras.
- feature2d: el cual incluye detectores de características, descriptores, etc.

Algunas funciones de los módulos de OpenCV anteriormente listados fueron utilizadas en la segmentación de imágenes para extraer formas y contornos de los objetos contenidos en las mismas, así como para comparar dichas formas o contornos. También se utilizaron funciones para leer, guardar y modificar imágenes, entre otras.

La publicación de OpenCV se da bajo la licencia BSD, lo cual permite que pueda ser utilizada libremente para propósitos de investigación y comerciales con las condiciones y restricciones que en dicha licencia se plantean.

#### 4.2.3. Alfred

Alfred (Bee y cols., 2009) es un sistema de animación facial capaz de crear expresiones para agentes animados de una manera fácil, intuitiva y rápida. Alfred utiliza el sistema de código

de acción facial FACS de Paul Ekman y Wallace Friesen para la descripción de expresiones faciales.

Algunas de las características de Alfred son:

- Un sistema de animación facial modularizado a través del sistema FACS lo cual permite generar una gran cantidad de expresiones faciales emocionales y conversacionales.
- Los ojos y cabeza son controlados por medio de un componente de cinemática inversa que le permite a Alfred mirar en cualquier dirección que se desee.
- Alfred es capaz de hablar y sincronizar los labios en tiempo real con archivos de audio pre-grabados.

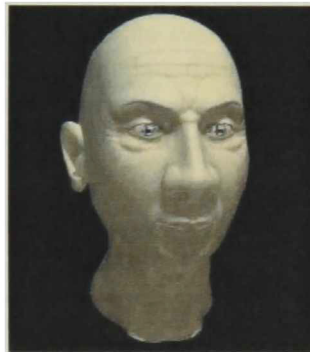


Figura 4.1: Alfred, herramienta utilizada para mostrar de manera explícita comportamientos ligados a la atención tales como movimientos de los ojos.

Alfred fue utilizado en este modelo para implementar la criatura o agente virtual cuya arquitectura subyacente incluye el modelo de atención propuesto en este trabajo. Basado en las capacidades de Alfred, la criatura virtual es capaz de simular cambios en la mirada (que implican movimientos de los ojos y cabeza) cuando se requiere la atención abierta (*overt attention*).

La distribución del código de Alfred es gratuita por lo que puede ser utilizado sin inconveniente alguno en esta implementación

### 4.3. Caso de Estudio: Inspección Libre (*Free-Viewing*)

El primer caso de estudio con el que se probará el funcionamiento del modelo de atención propuesto es el de *free-viewing* o inspección libre. Este tipo de prueba o experimento ha sido

utilizado para probar diversos modelos de atención visual entre los que se encuentran: (Itti y cols., 1998), (Itti y Koch, 2000), (Sun y cols., 2008), entre otros.

El experimento de inspección libre consiste en presentar una escena visual (generalmente una imagen 2D estática), en la cual se encuentran diversos objetos, y enseguida verificar el patrón de atención, es decir, cuales objetos o ubicaciones son atendidas en la escena. Este experimento se realiza tanto para verificar el comportamiento de personas como para sistemas artificiales de atención como el que se propone aquí. La mayoría de los sistemas de atención artificiales que han utilizado este experimento para probar su funcionamiento son sistemas de atención basados en ubicaciones espaciales, sin embargo este experimento también ha sido utilizado en sistemas de atención visual basados en objetos.

En base a la descripción general del experimento uno puede darse cuenta que no hay un objetivo explícito en la actividad; el objetivo implícito es inspeccionar, explorar, observar una imagen pero más allá de este objetivo no existe otro. Por lo tanto, en este experimento las influencias endógenas o *top-down* quedan de lado para probar básicamente como las influencias exógenas o *bottom-up* influyen en la atención. Puede decirse entonces que las plantillas de atención en este tipo de experimento no existen o son plantillas de atención vacías.

Dado que en este experimento los factores exógenos (físicos y semánticos) son los únicos que sesgan la competencia entre los estímulos, no se calculan valores de relevancia de los objetos en las imágenes sólo valores de prominencia de los mismos. En base a la prominencia (basada en características físicas y semántica) de los estímulos se decide que objeto de la imagen atender ya sea de manera abierta o cubierta, dependiendo de la ubicación del objeto que gane la competencia y la ubicación del área foveal en la imagen. Gracias al mecanismo IOR que se consideró en el modelo se puede atender diversos objetos (sin repetir alguno) y de esta forma obtener la trayectoria de atención o el conjunto de objetos que la criatura virtual atiende.

Si bien es cierto que muchas de las imágenes que se utilizan en este experimento son de escenas naturales, es decir, son imágenes de la vida cotidiana donde una gran variedad de objetos se encuentran presentes sin un acomodo estricto y donde muchas de las veces algunos objetos se encuentran parcialmente obstruidos por otros; en este caso de estudio nuestro objetivo fundamental no es probar el funcionamiento de este modelo con esta clase de imágenes sino más bien probarlo con imágenes que son utilizadas en experimentos de atención tanto psicológicos como neurocientíficos. Las imágenes que se utilizan en dichos experimentos son más sencillas ya que generalmente consisten de una matriz o arreglo de objetos perfectamente distinguibles y en donde uno o varios de ellos resaltan (*pop-out*) del resto. Al utilizar esta clase de imágenes, se tiene como objetivo el comprobar que el modelo propuesto es capaz de lidiar con imágenes sencillas y hasta cierto tipo estándares. Es decir, primero tenemos la intención de probar el funcionamiento del modelo con casos sencillos y

donde se conoce de antemano en muchos de las ocasiones el comportamiento que debería tener el modelo. Una vez que se comprueba que el modelo trabaja apropiadamente con estas entradas sencillas, entonces se verifica el funcionamiento del mismo con entradas más complejas en donde no es trivial conocer de antemano cuales objetos el modelo debería de atender.

### 4.3.1. Resultados

En este apartado se presentan los resultados que se obtuvieron en las diferentes simulaciones de los experimentos de inspección libre. Se consideraron dos tipos generales de inspección libre. En el primer tipo de inspección libre, la imagen que se presenta a la criatura virtual contiene un objeto que difiere del resto con respecto a sus características físicas (color, forma u orientación). En el segundo tipo de inspección libre, la imagen de entrada contiene un objeto que difiere del resto con respecto a su información semántica (congruencia semántica o familiaridad).

Las imágenes utilizadas en este caso de estudio contienen un número de objetos que varía entre 6 y 36 objetos por imagen. En la figura 4.2 se muestran algunas imágenes que fueron utilizadas en las simulaciones del primer tipo de inspección libre.

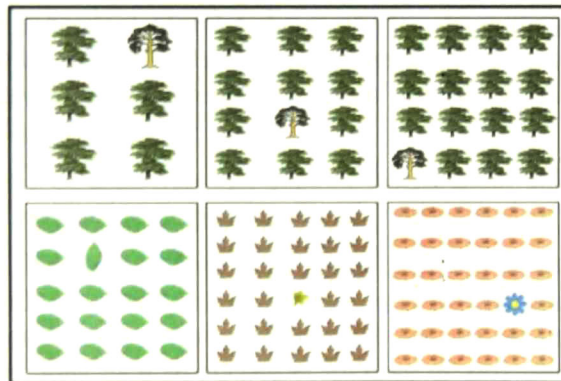


Figura 4.2: Imágenes utilizadas para el primer tipo de inspección libre.

Por otro lado, la figura 4.3 muestra algunas imágenes de entrada que fueron utilizadas en las simulaciones del segundo tipo de inspección libre.

Antes de presentar los resultados que arrojaron las simulaciones de los dos tipos de inspección libre, se describirá de manera breve como se lleva a cabo la ejecución de las mismas. La ejecución comienza presentando en una misma ventana a la criatura virtual, representada

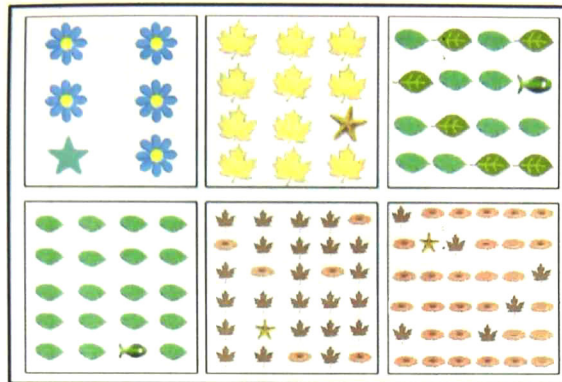


Figura 4.3: Imágenes utilizadas para el segundo tipo de inspección libre.

por Alfred, junto con la imagen de entrada la cual representa el ambiente percibido por la criatura. Una vez que la imagen es capturada se realiza toda la serie de operaciones del modelo que simulan la extracción de características de la imagen y la competencia de los objetos sesgada sólo por influencias exógenas en este caso. Enseguida se determina el objeto ganador de la competencia y se procede a atender dicho objeto, dependiendo del centro de mirada de la criatura virtual y la ubicación del objeto a atender en la imagen este se atenderá de manera abierta o cubierta. Si se atiende de manera abierta, Alfred moverá sus ojos en la dirección de dicho objeto.

En las figuras 4.4 y 4.5 se muestran ejemplos de la ejecución del modelo con imágenes de entrada para el caso de estudio de inspección libre.

Una de las maneras de medir el desempeño de un modelo de atención visual consiste en determinar que tan rápido se atiende a un estímulo que sobresale del resto. En este modelo no se mide el tiempo que tarda en atenderse dicho objeto en unidades de tiempo como milisegundos o segundos, más bien se cuenta el número de objetos que son atendidos antes de que el objeto que destaca en la imagen es atendido, es decir, el número de detecciones fallidas. Cabe señalar que todas las imágenes utilizadas en los experimentos de inspección libre constan de un objeto que sobresale del resto (física o semánticamente según el tipo) y un conjunto de distractores u objetos que no resaltan en demasía.

La figura 4.6 muestra los resultados obtenidos en las simulaciones de los dos tipos de inspección visual considerando el número de detecciones fallidas con respecto a la cantidad de distractores contenidos en la imágenes de entrada. En la última sección de este capítulo se presentan algunas conclusiones hechas a la luz de los resultados arrojados por estas simulaciones.

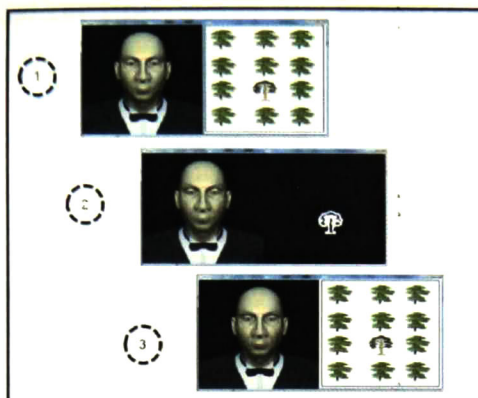


Figura 4.4: Ejecución del modelo de atención para el primer tipo de inspección libre. La imagen de entrada contiene un objeto que difiere del resto en sus características físicas, principalmente. En este ejemplo, el objeto que gana la competencia por atención (árbol de color café) se encuentra prácticamente en centro de la mirada de Alfred por lo que es atendido de manera cubierta.

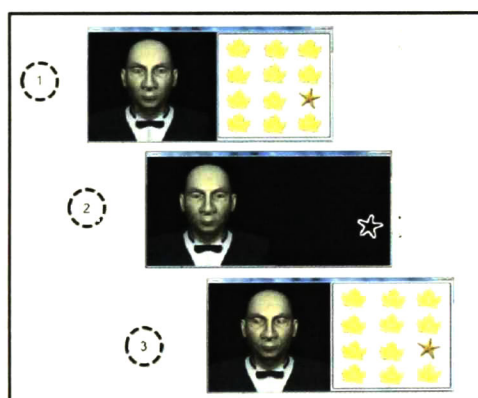


Figura 4.5: Ejecución del modelo de atención para el segundo tipo de inspección libre. La imagen de entrada contiene un objeto que difiere del resto con respecto a su información semántica. En este ejemplo, el objeto que gana la competencia (estrella de mar de color naranja) se encuentra fuera del centro de la mirada de Alfred, por lo tanto se realiza un movimiento de los ojos para atenderlo.

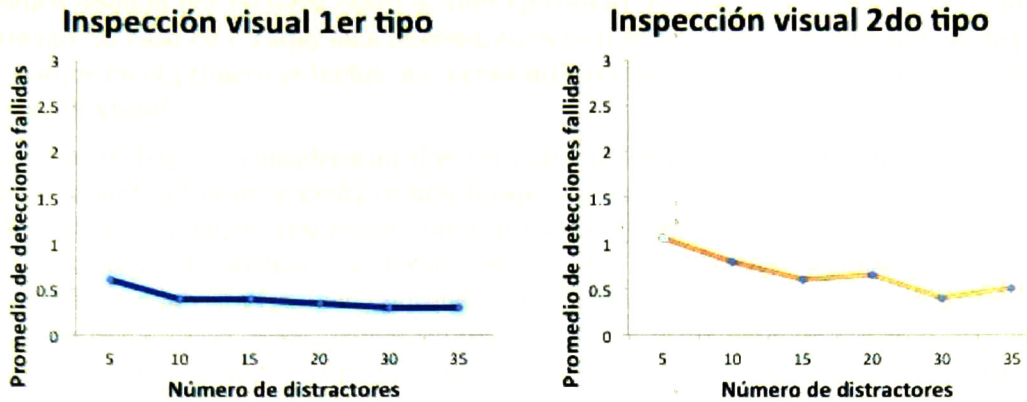


Figura 4.6: Resultados de las simulaciones de los dos tipos de inspección visual.

#### 4.4. Caso de Estudio: Búsqueda Visual (*Visual Search*)

El segundo caso que se consideró para probar el funcionamiento del modelo de atención propuesto es el de búsqueda visual (*visual search*), el cual es muy común en el campo de la atención visual. Este experimento y sus variaciones ha sido utilizado en diversos trabajos tales como (Li, 2002), (Bisley y cols., 2011), por citar algunos.

El experimento de búsqueda visual consiste básicamente en presentar a un individuo una imagen en la cual el sujeto debe encontrar cierto objetivo previamente establecido. Es decir, primero se presenta (por medio de una imagen) el objetivo a buscar y enseguida se procede a presentar la imagen en la cual este objetivo deberá ser buscado. Existen diversas variaciones de este experimento donde se cambia la complejidad de la imagen y las características del objetivo a buscar.

Considerando la descripción general del experimento de búsqueda visual uno puede notar que este experimento, a diferencia del experimento de inspección visual, incorpora no sólo influencias exógenas sino también influencias endógenas en la competición de los estímulos por atención. Dado que en la búsqueda visual se tiene un objetivo explícito (un objeto con determinadas características a buscar), entonces en este experimento se generará una plantilla de atención que representará dicho objetivo. Utilizando la información de la plantilla de atención, en el modelo propuesto se calcularán valores de relevancia (basada en características físicas y semántica) para los objetos contenidos en la imagen. Cabe señalar que también se calcularán valores de prominencia que junto con las relevancias darán lugar al cálculo de prioridad de los objetos. Es decir, en este caso se podrá observar una competencia entre los

estímulos sesgada por factores tanto *bottom-up* como *top-down*. Sin duda la búsqueda visual representa un caso de estudio más interesante pero a su vez más complejo que la inspección libre porque en el primero se incluyen muchas más variables y factores en la competencia por la atención visual.

En este trabajo se consideraron dos variantes del experimento de búsqueda visual. La primera variante (*feature search*) es una búsqueda visual donde el objetivo difiere del resto de objetos en la imagen, conocidos como distractores, en una sola característica física tal como el color, la orientación o la forma. Se dice que en este tipo de búsqueda el objetivo destaca o salta a la vista (*pop-out*) porque al no tener atributos en común con los distractores su característica única sufre menos supresión que las otras características o atributos de la imagen. Generalmente, el tiempo que lleva el encontrar el objetivo en esta variante de búsqueda visual es relativamente corto. En la segunda variante del experimento de búsqueda visual (*conjunction search*), el objetivo no se define por un único atributo visual sino por una combinación de dos o más características. Este tipo de búsqueda aunque si bien sencilla, tiene más complejidad que la primera variante. Se sabe que el tiempo empleado para encontrar el objetivo en la imagen es mayor que el de la primera variante debido a que debe integrarse información de dos o más atributos visuales con el fin de localizar el objetivo.

La finalidad de llevar a cabo el experimento de búsqueda visual, con sus dos variantes, es la de probar el funcionamiento del modelo de atención propuesto con influencias exógenas y endógenas. También debemos ser capaces de replicar en cierta medida los resultados que se obtienen en las dos variantes, es decir, que cuando el objetivo es más complejo (con respecto a características físicas y semánticas) el tiempo para encontrarlo se incrementa. Al igual que en el caso de inspección visual, las imágenes utilizadas en este experimento no son imágenes de escenas naturales sino más bien imágenes similares a las utilizadas en experimentos psicológicos y neurocientíficos.

#### 4.4.1. Resultados

En esta sección se presentan los resultados obtenidos en las diferentes simulaciones de los experimentos de búsqueda visual tanto de *feature search* como de *conjunction search*.

Cabe señalar que las imágenes utilizadas en este caso de estudio, en los dos tipos de variante, fueron tomadas de la base de datos de imágenes que se utilizó para la ejecución del modelo de Itti y Koch (2000).

Las imágenes utilizadas en ambos experimentos de búsqueda visual contienen, al igual que en los experimentos de inspección libre, un número de objetos cuya cantidad varía entre 6 y 36 objetos por imagen. En la figura 4.7 se muestran algunas imágenes que fueron utilizadas para el experimento de búsqueda visual en la variante *feature-search*.



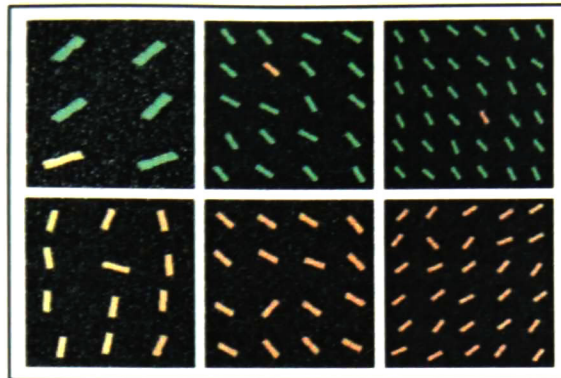


Figura 4.7: Imágenes utilizadas en la variante *feature search* de búsqueda visual.

En la figura 4.8 se muestran algunas imágenes que fueron utilizadas para el experimento de búsqueda visual en la variante *conjunction-search*.

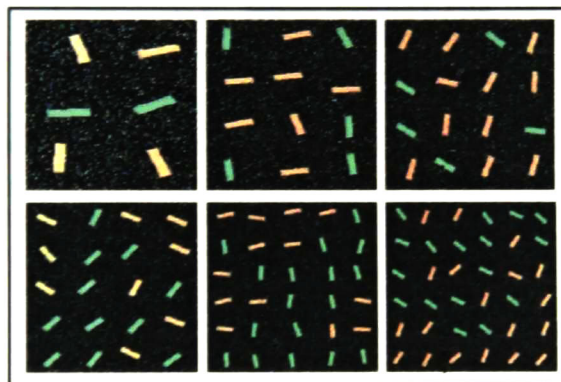


Figura 4.8: Imágenes utilizadas en la variante *conjunction search* de búsqueda visual.

En el caso de estudio de inspección libre se explicó brevemente como se da la ejecución del modelo propuesto. Para este caso se procede de manera igual que el caso anterior, la diferencia radica en que en el caso de búsqueda visual si se incluyen influencias endógenas. Antes de presentar la imagen de entrada se crea una plantilla en base al objeto a buscar y la información de esta se envía a las diferentes áreas de procesamiento en la vía ventral para influenciar la competencia por atención. Después se presenta la imagen como en el caso de estudio anterior y se llevan a cabo los mecanismos para simular la competencia entre los objetos de la imagen y se determina el objeto a atender y la manera en que este será atendido

(*covert attention* o *overt attention*). Una vez que el objeto es atendido, se determina si es el objeto que se está buscando. Si lo es, termina la ejecución de ese experimento, si no la ejecución continua hasta que se encuentre el objeto deseado.

En las figuras 4.9 y 4.10 se muestran ejemplos de la ejecución del modelo con imágenes de entrada para el caso de estudio de búsqueda visual en la variante *feature-search* y *conjunction search*, respectivamente.

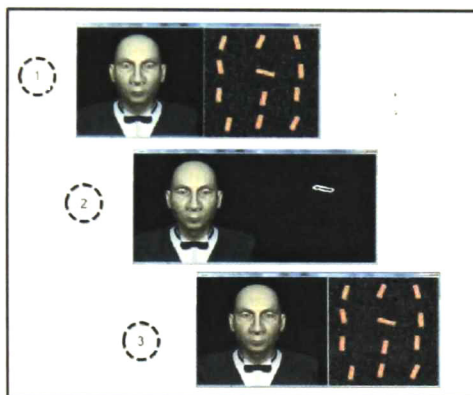


Figura 4.9: Ejecución del modelo de atención para el caso de estudio de búsqueda visual en la variante *feature search*. En este ejemplo, dado que el objeto a buscar (barra horizontal roja) se encuentra ligeramente arriba del centro de la mirada de Alfred, se requirió de un movimiento ocular para atenderlo y posteriormente determinar que es el objeto que se desea encontrar.

En los experimentos de búsqueda visual comúnmente se mide el tiempo necesario para encontrar el objetivo en la imagen de entrada. En los experimentos realizados para este modelo de atención no se mide el tiempo en segundos o en unidades de tiempo estándares (segundos, décimas de segundo, etc.). El tiempo que se tarda en encontrar un objeto se mide en base a las detecciones fallidas como las que se explicaron en el caso de inspección libre, es decir, el número de objetos que se atendió antes de encontrar el objetivo. Las imágenes utilizadas en estas simulaciones constan de un solo objetivo y un conjunto de objetos distractores irrelevantes.

En la figura 4.11 se presentan los resultados arrojados por las simulaciones de los experimentos de búsqueda visual considerando el número de detecciones fallidas con respecto a la cantidad de distractores contenidos en la imágenes. Las conclusiones realizadas en base a estos resultados se presentan en la siguiente sección.

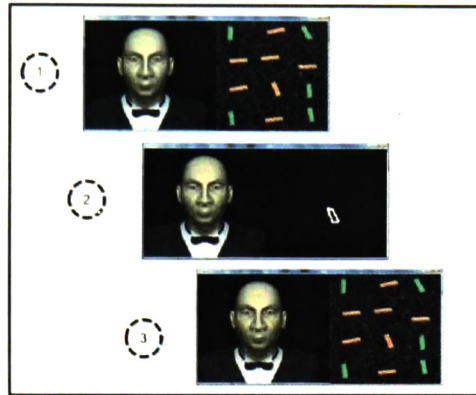


Figura 4.10: Ejecución del modelo de atención para el caso de estudio de búsqueda visual en la variante *conjunction search*. En este ejemplo, debido a que el objeto a buscar (barra roja vertical) se encuentra prácticamente en centro de la mirada de Alfred, no fue necesario ningún movimiento de los ojos para atenderlo y determinar que ha sido encontrado.

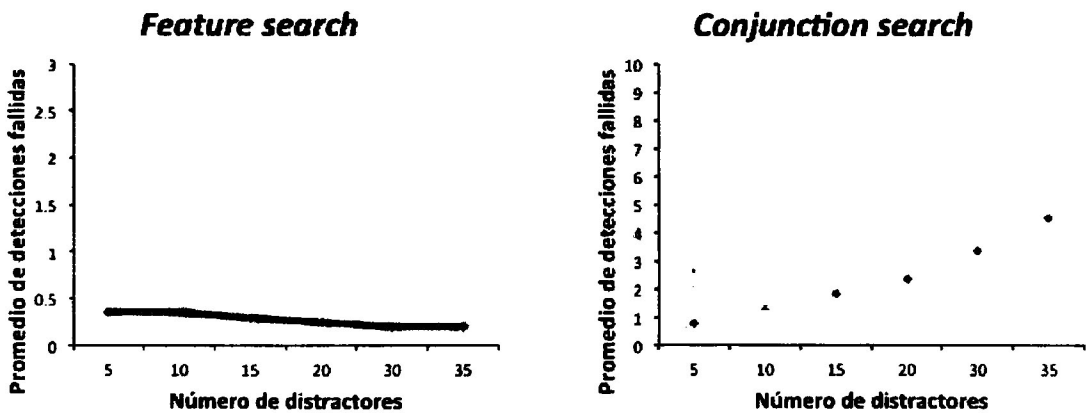


Figura 4.11: Resultados de los experimentos del caso de estudio de búsqueda visual.

## 4.5. Conclusiones

A partir de los resultados obtenidos en las simulaciones del caso de estudio de inspección visual se pueden realizar varias conclusiones. La primera de ellas es que el modelo de atención visual que se propone tiene la capacidad de atender de manera rápida aquellos objetos que resaltan o destacan (físicamente y semánticamente) en una escena visual. Cuando se habla de rapidez en este contexto, se refiere a que el número de detecciones fallidas o el número de objetos que se atienden antes de atender al objeto más destacado en la imagen es muy pequeño. Al observar la figura 4.6 se puede ver que para ambos tipos de inspección visual se tiene un patrón descendiente del promedio de detecciones fallidas conforme aumenta el número de distractores en la imagen. Lo anterior podría explicarse en términos del espacio que existe entre los objetos de la imagen. Dado que las imágenes utilizadas son prácticamente del mismo tamaño, independientemente del número de objetos que contengan, entonces cuando se tiene un número considerable de objetos el espacio entre estos es muy poco. Esto provocaría que el objeto más destacado se atendiera más rápido debido a que la prominencia de este no sería absorbida ni por el fondo ni los distractores alrededor. En cambio, cuando se tienen menos objetos en la imagen estos son de mayor tamaño y el espacio entre estos es menor por lo que la prominencia del objeto más destacado en ocasiones podría ser absorbida por el fondo de la imagen y los objetos a su alrededor. Sin embargo esto es sólo una teoría, como puede observarse en la gráfica, no es mucha la diferencia entre el promedio de detecciones fallidas entre el caso con menor número de distractores y el que tiene mayor cantidad de estos.

La gráfica de inspección visual del segundo tipo muestra la rapidez del modelo cuando se tiene un objeto prominente en términos de semántica. Sin embargo, el promedio de detecciones fallidas es mayor que el del otro tipo de inspección. La prominencia semántica de los objetos en el modelo propuesto compite con la prominencia física para determinar que objeto atender, por lo tanto el promedio ligeramente mayor en este caso podría deberse a que si los distractores son prominentes en términos de atributos físicos, la atención podría sesgarse a unos de estos antes de atender al objeto que más destaca con respecto a la información semántica.

Con respecto a los resultados del caso de estudio de búsqueda visual se concluye que el modelo provee un mecanismo eficaz para buscar un objetivo en una imagen dada. La eficacia se observa en la cantidad de detecciones fallidas que se realizan antes de encontrar el objeto deseado. En el caso de *feature search* puede decirse que la detección del objetivo es casi automática independientemente del número de distractores. Cuando el estímulo a buscar destaca con respecto del resto es más rápido encontrarlo. La prominencia de este objeto destacado y su relevancia en la tarea actual, provocan que su detección sea casi instantánea lo cual es congruente con lo que ocurre en las personas (Desimone y Duncan, 1995). Cuando se hace una búsqueda visual de tipo *conjunction search*, donde los distractores tienen uno o

varios atributos en común con el objetivo, el tiempo requerido para encontrar dicho objetivo es mucho mayor comparado con el requerido en *feature search* (Li, 2002). Este comportamiento que se presenta en los seres humanos se ve reflejado en este modelo ya que el número de detecciones fallidas es mayor que el de *feature search*. En el caso de *conjunction search*, entre más distractores se tengan en la imagen, más tiempo se requiere para encontrar el objetivo lo cual coincide con el comportamiento mostrado por los seres humanos. Si bien es cierto que el promedio de detecciones fallidas se incrementa en el caso de búsqueda visual de tipo *conjunction search*, la relevancia del objeto a buscar (expresada en términos de plantillas de atención) provoca que el número de detecciones fallidas no se dispare como podría esperarse (Itti y Koch, 2000), es decir, la relevancia o influencia endógena favorece el rendimiento en este tipo de experimento.

En resumen podemos decir que el modelo propuesto provee un mecanismo rápido para atender objetos que resaltan a la vista tanto por características físicas y semánticas lo cual es fundamental ya que las personas estamos constantemente sesgadas a atender este tipo de objetos. También el modelo provee un mecanismo eficaz de búsqueda visual sobre todo cuando se tienen casos difíciles donde el objeto a encontrar no resalta mucho con respecto a los distractores. Esto gracias a que no solo se consideran atributos físicos sino la relevancia de los objetos en la tarea actual para la competencia de atención.

# Capítulo 5

## Conclusiones

En el primer capítulo de este documento se delinearon los objetivos que se persiguen en esta trabajo. Partiendo de estos objetivos y del análisis de diversas aproximaciones de modelos de atención desarrollados previamente, se presentó un modelo de atención visual para criaturas virtuales en el que diferentes factores y mecanismos intervienen. Posteriormente, se mostraron los resultados arrojados por las simulaciones realizadas para probar el funcionamiento del modelo propuesto.

En este capítulo se presentan las conclusiones de este trabajo las cuales se realizaron en base a los objetivos, el diseño del modelo y los resultados arrojados por las simulaciones realizadas para probar su funcionamiento.

### 5.1. Conclusiones

Las conclusiones que se pueden hacer acerca de cualquier modelo consideran el o los problemas que se intentan atacar con dicho modelo, los objetivos que se persiguen, el modelo como tal y los resultados que produce la implementación de dicho modelo. En el primer capítulo de este documento se definió como problema el poder dotar a criaturas virtuales con mecanismos que les permitan exhibir comportamientos más creíbles. Un proceso cognitivo que permitirá a dichas criaturas el tener comportamientos más reales es el proceso de atención, específicamente la atención visual. Este proceso permitirá a las criaturas virtuales concentrarse sólo en aquella información destacada o relevante de entre toda la información que continuamente están recibiendo tal y como sucede con los seres humanos.

La atención está influenciada por diversos factores que son determinantes a la hora de seleccionar la información que será atendida en un momento dado. Información proveniente de otros procesos cognitivos tales como memoria y emociones pudiera influir en el proceso de atención. Considerando esto y algunas aproximaciones previas que intentan modelar el

proceso de atención desde diversas perspectivas, se diseñó un modelo de atención que pudiera atacar de una mejor manera la problemática identificada.

El modelo de atención visual que se propone en este trabajo considera los dos tipos de influencias principales que se han identificado como determinantes en el proceso de atención, es decir, influencias exógenas y endógenas. A través de una competencia entre los estímulos sesgada por estas influencias se decide que objeto atender. El hecho de incluir los dos tipos de influencias en el modelo de atención es importante ya que se acerca un poco más a como se da dicho proceso en los seres humanos. Entre estas influencias, particularmente entre las influencias exógenas, en este modelo se contemplaron influencias provenientes de otros procesos cognitivos: memoria y emociones, específicamente. La influencia de otros procesos cognitivos en el proceso de atención ha sido ignorada en aproximaciones previas, sin embargo en este modelo no ha sido así y esta es una de las ventajas que se tiene con respecto a otras aproximaciones. El incluir influencias de otros procesos cognitivos en la atención no ha sido un mero capricho o una ocurrencia sin sustento si no que ha sido inspirado en base a evidencia científica. Esto da como resultado un modelo que no considera a la atención como una función aislada sino como un proceso en el que diferentes mecanismos y procesos cognitivos tienen ingerencia.

Al considerar evidencia de neurociencia para el diseño de este modelo, específicamente las áreas cerebrales que se encuentran involucradas en la atención y el papel de estas, se pretende modelar de manera un poco más detallada como la atención se lleva a cabo. No se pretende modelar el funcionamiento exacto de estas áreas ni mucho menos modelar el funcionamiento de un cerebro humano de manera fidedigna, simplemente se desea que este modelo sea parte de una arquitectura que pueda permitir a criaturas virtuales comportamientos más creíbles. Por ello se consideró que el utilizar dicha evidencia ayudaría a entender el funcionamiento de la atención de una manera más certera y por ende crear un modelo que genere comportamientos más parecidos a los que se observan en los humanos.

El hecho de que el modelo presentado en este trabajo sólo trabaja con información visual no significa que este modelo no sea útil. Los resultados arrojados por las simulaciones de dos casos de estudio (con sus respectivas variantes) demostraron la utilidad de este modelo. Si bien estos casos de estudio podrían considerarse poco complejos no significa que el modelo propuesto sólo pueda trabajar con este tipo de casos. Simplemente se intentó mostrar que el modelo cumple con los requisitos primarios con los que un modelo de atención debe cumplir. A partir de esto, entonces podrían realizarse experimentos más complejos considerando procesos adicionales o diversos tipos de información y sesgos.

Otra ventaja del modelo propuesto es que al modelarse la atención como un proceso distribuido entre varios componentes, el extender el modelo o el añadir nuevas influencias en él es más natural y menos conflictivo.

Con respecto a las debilidades del modelo podemos considerar el que se trabaja con imáge-

nes 2D y no con secuencias de video, por ejemplo. Sin embargo, el modelo no está cerrado sólo a esta posibilidad ya que una secuencia de video es una serie de imagenes presentadas de manera continua. Otra debilidad del modelo pudiera ser el hecho de que sólo se modeló un tipo de influencia endógena, los objetivos, y de manera sencilla. Aspectos como las intenciones, estado interno, deseos y conocimiento no fueron considerados ya que son bastante complejos de modelar. A pesar de esto consideramos útil el haber utilizado plantillas de atención para representar los objetivos ya que podrían considerarse plantillas o elementos análogos para representar otras influencias endógenas. Aunque el modelo sólo incluye información visual, esto no indica que no pueda incluirse información de otro tipo de modalidad sensorial ya que las influencias exógenas y endógenas aplican para las otras modalidades sensoriales y no sólo para las visuales.

En resumen se puede concluir que el modelo de atención visual presentado en este trabajo representa un buen comienzo en el intento por modelar la atención no como un proceso aislado sino como un proceso que influye y es influenciado por el funcionamiento de otros procesos cognitivos concurrentes. Al modelarse la atención de esta forma se acerca un poco más a la forma como se da esta en las personas. Por lo tanto, el dotar a criaturas virtuales con este modelo les permitirá exhibir comportamientos más creíbles. Aunque los comportamientos generados por este modelo de atención pudieran considerarse sencillos, el modelo propuesto puede considerarse como una base para exhibir comportamientos mucho más complejos en los que se considera una gama más amplia de factores e influencias.

## 5.2. Trabajo Futuro

En este trabajo se modeló la atención como un proceso cuya función principal es filtrar la información que compite por ingresar a la memoria de trabajo. Sin embargo, la atención también contempla aspectos de vigilancia, es decir, cuando y durante cuanto tiempo se atenderá a un determinado estímulo. Otro aspecto que se contempla en la atención es el control, es decir, decidir cuando atender un estímulo y cuando hacer el cambio de atención hacia otro estímulo. Sin duda estos aspectos son de un alto nivel de cognición por lo que modelarlos requiere de un estudio más profundo de ellos. Entonces, el estudio de estos aspectos y su modelado e inclusión en el modelo de atención puede considerarse como trabajo futuro necesario si es que se quiere utilizar el modelo para generar comportamientos más complejos.

Un aspecto que en este modelo no se consideró es el de la atención dividida. En la atención dividida se puede atender a más de un objeto o estímulo. Esta propiedad la exhibimos los seres humanos, sin embargo nuestra capacidad de atender es limitada por lo que atender a más de un estímulo disminuye sensiblemente esta capacidad. Por ejemplo hablar por celular y conducir un vehículo es posible pero el hacer esto disminuye seriamente el desempeño en alguna o en ambas actividades. Dotar al modelo con esta característica no es sencillo ya que



se tienen implícitos los aspectos de la atención mencionados atrás: vigilancia y control.

Otro aspecto en el que podría ahondarse en este modelo es en el modelado de los otros tipos de influencias endógenas tales como expectativas, conocimiento, etc., que sin lugar a duda potenciaría las capacidades del modelo propuesto y su vez el de las criaturas virtuales que cuenten con él.

Aunque el sentido de la vista es el más estudiado por la comunidad científica, esto no indica que el resto de sentidos y la información recibida a través de estos no es importante. En el cerebro humano la información visual no es el único tipo de información que compite por procesamiento. Información de otras modalidades sensoriales también compite por capturar la atención. Incluir diversos tipos de información sensorial en el modelo de atención ampliará el conjunto de aplicaciones donde podrían utilizarse las criaturas virtuales dotadas con dicho modelo.

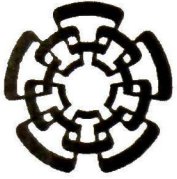
Como puede notarse, la atención engloba un gran número de aspectos que el modelo propuesto no cubre de manera total, sin embargo con lo expuesto en esta sección se tiene cierta certeza acerca de cual serían los siguientes pasos con el fin de extender el modelo de atención para que pueda ser utilizado en situaciones y experimentos más complejos.

# Referencias

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., y Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.
- Anderson, J. R., Matessa, M., y Douglass, S. (1995). The ACT-R theory and visual attention. En *Proceedings of the seventeenth annual conference of the cognitive sciences society* (pp. 61–65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baluch, F., y Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, *34*(4), 210–224.
- Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.*, *5*, 617–629.
- Bee, N., Falk, B., y André, E. (2009). Simplified facial animation control utilizing novel input devices: A comparative study. En *International Conference on Intelligent User Interfaces (IUI '09)* (pp. 197–206).
- Bisley, J. W., Mirpour, K., Arcizet, F., y Ong, W. S. (2011). The role of the lateral intraparietal area in orienting attention and its implications for visual search. *European Journal of Neuroscience*, *33*, 1982–1990.
- Deco, G., y Zihl, J. (2001). A neurodynamical model of visual attention: Feedback enhancement of spatial resolution in a hierarchical system. *Journal of Computational Neuroscience*, *10*, 231–253.
- Desimone, R., y Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.
- Engel, S., Zhang, X., y Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, *388*(6637), 68–71.
- Eriksen, C. W., y James, J. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, *40*(4), 225–240.
- Fecteau, J. H., y Munoz, D. P. (2006). Saliency, relevance and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390.
- Itti, L., y Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.
- Itti, L., Koch, C., y Niebur, E. (1998, November). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(11), 1254–1259.

- James, W. (1890). *The principles of psychology* (Vol. 1). New York: Henry Holt.
- Kastner, S., y Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.*, *23*, 315–341.
- Kiani, R., Esteky, H., Mirpour, K., y Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*, 4296–4309.
- Knudsen, E. I. (2007). Fundamental components of attention. *Annu. Rev. Neurosci.*, *30*, 57–78.
- Koch, C., y Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.
- Laird, J. E., y Congdon, C. B. (2012, April). The Soar's user manual [Manual de software informático].
- Leventhal, A. G. (1991). *The neural basis of visual function: Vision and visual dysfunction*. CRC Press.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16.
- Lund, J. S., y Boothe, R. G. (1975). Interlaminar connections and pyramidal neuron organization in the visual cortex, area 17, of the macaque monkey. *Journal of Comparative Neurology*, *159*, 305–334.
- Machado, H., y Rafal, R. (2000). Strategic control over saccadic eye movements: Studies of the fixation offset effect. *Attention, Perception, and Psychophysics*, *62*(6), 1236–1242.
- Martinho, C., y Ana, P. (2008). Anticipation and believability. En G. Pezzulo, M. Butz, C. Castelfranchi, y R. Falconi (Eds.), *The Challenge of Anticipation* (Vol. 5255, p. 115-133). Springer Berlin / Heidelberg.
- Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annu. Rev. Neurosci.*, *16*, 245–263.
- Öhman, A., Flykt, A., y Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology*, *130*(3), 466–478.
- Purves, D., y cols. (Eds.). (2001). *Neuroscience* (2nd Edition ed.). Sinauer Associates.
- Ruesch, J., Lopes, M., Bernardino, A., Hörnstein, J., Santos-Victor, J., y Pfeifer, R. (2008). Multimodal saliency-based bottom-up attention- a framework for the humanoid robot iCub. En *Proceedings of IEEE International Conference on Robotics and Automation* (pp. 962–967). Pasadena, CA, USA.
- Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences*, *8*(5), 224–230.
- Sun, Y., Fisher, R., Wang, F., y Gomes, H. M. (2008). A computer vision model for visual-object-based attention and eye movements. *Computer Vision and Image Understanding*, *112*(2), 126–142.
- Underwood, G., y Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Quarterly journal of Experimental Psychology*,

- 59(1931–1949).
- Ungerleider, L. G., y Mishkin, M. (1982). Two cortical visual systems. En J. Ingle, M. A. Goodale, y R. J. W. Mansfield (Eds.), *The Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: The MIT Press.
- VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology (Paris)*, 97, 365–377.
- Vernon, D., Metta, G., y Sandini, G. (2007, July). The iCub cognitive architecture: Interactive development in a humanoid robot. En *IEEE 6th International Conference on Development and Learning* (pp. 122–127).
- Võ, M. L.-H., y Henderson, J. M. (2009). Does gravity matter? effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3:24), 1–15.
- Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, 9(12), 585–594.
- Weismeyer, M., y Laird, J. (1993). NOVA, covert attention explored through unified theories of cognition. En *Proceedings of the fifteenth annual conference of the cognitive science society* (pp. 102–107).
- Werblin, F., Roska, B., y Balya, D. (2001). Parallel processing in the mammalian retina: lateral and vertical interactions across stacked representations. *Progress in Brain Research*, 131(229–238).
- Wurtz, R. H., y Kandel, E. R. (2000). *Principles of Neural Science* (4th ed.). Mc Graw Hill.
- Young, R. M., y Lewis, R. L. (1999). The Soar cognitive architecture and human working memory. En A. Miyake y P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 224–256). New York: Cambridge University Press.



# **CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL I.P.N UNIDAD GUADALAJARA**

El Jurado designado por la Unidad Guadalajara del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional aprobó la tesis

**Modelo del Sistema Cognitivo de Atención para Criaturas Virtuales**

del (la) C.

**Silviano DÍAZ BARRIGA SOSA**

el día 10 de Agosto de 2012.

**Dr. Luis Ernesto López Mellado**  
Investigador CINVESTAV 3C  
CINVESTAV Unidad Guadalajara

**Dr. Juan Luis Del Valle Padilla**  
Investigador CINVESTAV 3C  
CINVESTAV Unidad Guadalajara

**Dr. Félix Francisco Ramos Corchado**  
Investigador CINVESTAV 3A  
CINVESTAV Unidad Guadalajara

**Dr. Marco Antonio Ramos Corchado**  
Profesor Investigador Nivel F  
Universidad Autónoma del Estado de  
México

**Dr. Pablo Covarrubias Salcido**  
Profesor Investigador  
Universidad de Guadalajara



CINVESTAV - IPN  
Biblioteca Central



SSIT0011081