



Facultad de Psicología

Trabajo de  
fin de grado

**Modalidad 2**

**“Proyecto de investigación empírica  
y/o desarrollo de un estudio piloto”**

Significatividad de la  
red conceptual de una  
imagen en condiciones  
de atención limitada

Análisis en función de los  
componentes del *gist*

Autor del TFG

David Gallego Conde

**Grado en Psicología**

Año 2019

# Índice

<b>Índice</b> .....	<b>2</b>
<b>Resumen</b> .....	<b>4</b>
<b>Abstract</b> .....	<b>5</b>
<b>Introducción</b> .....	<b>6</b>
1. ¿Cómo interpretamos una imagen? .....	6
1.1 Los componentes del <i>gist</i> .....	8
1.1.1 Propiedades estadísticas .....	8
1.1.2 Información estructural .....	9
1.1.3 Objetos.....	9
2. Atención visual y <i>gist</i> .....	10
2.1 La atención visual y los componentes del <i>gist</i> .....	11
3. El recuerdo del <i>gist</i> .....	12
<b>Experimento 1</b> .....	<b>14</b>
1. Método .....	15
1.1 Participantes.....	15
1.2 Materiales y Procedimiento .....	15
1.3 Análisis estadísticos.....	18

2.	Resultados .....	20
3.	Discusión.....	22
<b>Experimento 2 .....</b>	<b>23</b>	
1.	Método .....	24
1.1	Participantes.....	24
1.2	Materiales y Procedimiento .....	24
1.3	Análisis estadísticos.....	25
2.	Resultados .....	26
3.	Discusión.....	28
3.1	Comparación entre experimentos .....	28
3.1.1	Recuerdo de estímulos completos .....	28
3.1.2	Recuerdo de estímulos fragmentados .....	29
<b>Discusión general.....</b>	<b>31</b>	
<b>Conclusiones .....</b>	<b>34</b>	
<b>Referencias bibliográficas .....</b>	<b>35</b>	
<b>Índice de figuras .....</b>	<b>38</b>	
<b>Índice de tablas.....</b>	<b>39</b>	

## Resumen

Un vistazo rápido a una imagen es suficiente para que un observador pueda extraer la información necesaria para entenderla, identificarla e, incluso, recordarla; esto es, el *gist*. Sin embargo, esto resulta más complicado cuando la atención no está dirigida a la escena, sino a la realización de una tarea concurrente. En este estudio se analiza en qué medida los componentes del *gist* de una imagen (propiedades estadísticas, información estructural e información de objetos; Wolfe et al., 2007) pueden ser procesados y recordados en condiciones de atención limitada.

Para manipular los recursos atencionales dedicados a la imagen se diseñaron dos tareas: a) una de memoria, en la que los participantes únicamente debían observar las imágenes y b) una tarea dual, en la que debían realizar también una búsqueda visual. Además, se han manipulado los componentes del *gist*, bien fragmentando las imágenes para comprometer la información estructural, o bien utilizando texturas para restringir la información procedente de los objetos. Posteriormente, se evaluó el recuerdo mediante una tarea de reconocimiento.

La peor ejecución se dio en las condiciones de tarea dual, lo que muestra la necesidad de la atención para procesar correctamente el *gist*. Además, la información estructural resultó ser el componente más destacado en el reconocimiento de imágenes en condiciones de atención limitada. La memoria conceptual a corto plazo (Potter, 1976) puede ayudarnos a interpretar estos resultados: tanto la falta de información estructural, como la de objetos, limitan la significatividad de la red conceptual que se crea a partir de la imagen al no permitir el establecimiento de nexos con la información contenida en la memoria a largo plazo, por lo que el *gist* decae rápidamente.

**Palabras clave:** memoria de imágenes, atención, tarea dual, *gist*, procesamiento de escenas.

**Número de palabras del trabajo:** 7373

## Abstract

A quick glance at an image is enough for an observer to extract the information needed to understand it, identify it and even remember it: the gist. However, this becomes more complicated when the attention is not focused on the scene, but on the execution of a concurrent task. This study analyses to what extent the components of the gist of an image (statistical properties, structural information and object information; Wolfe et al., 2007) can be processed and remembered under conditions of limited attention.

Two tasks were designed in order to manipulate the attentional resources intended to engage the image: a) a memory task, in which participants only had to observe images and b) a dual task, in which they also had to perform a visual search. Gist components have also been manipulated, shuffling images to corrupt the structural information or using textures to restrict object-related information. Subsequently, memory was evaluated through a recognition task.

The worst performance took place in dual task conditions. This shows how necessary attention is to correctly process the gist. In addition, structural information turned out the most prominent component in image recognition under limited attention conditions. Conceptual short-term memory (Potter, 1976) can help us interpret these results, since both the lack of structural information and the lack of objects limit the significance of the conceptual network created from an image by not allowing the establishment of links with information included in the long-term memory. As a result, the gist rapidly decays.

**Keywords:** picture memory, attention, dual-task, gist, scene processing

# Introducción

## 1. ¿Cómo interpretamos una imagen?

Cada día, nuestro sistema visual recibe miles de imágenes provenientes del exterior, muchas de ellas de una duración tan breve como una sacada pero que resulta suficiente para entenderlas. Imaginemos que nos encontramos a oscuras en una sala cerrada y, de repente, alguien toma una foto utilizando el flash. A pesar de la breve duración del destello de luz, somos capaces de entender que nos encontramos en una cocina grande y alargada, que los muebles son de madera antigua, que la encimera recorre la pared de nuestra derecha y hace esquina con la pared de piedra del fondo, e incluso que un cajón está abierto; información suficiente para darnos cuenta de que nunca hemos estado en esa cocina. Ahora imaginemos que salimos de esa estancia, y la persona que ha tomado la foto nos la enseña, ¿seríamos capaces de reconocer que esa es la sala en la que acabamos de estar? En caso afirmativo, ¿qué información nos permitiría distinguir esa cocina de otra con las mismas características? ¿el recuerdo de la cocina sería peor si en el momento en el que se disparó el flash estuviésemos buscando un interruptor de luz en la pared?

Los observadores somos capaces de identificar una imagen extraordinariamente rápido (en aproximadamente 100 ms), pero necesitamos unos 300 ms de procesamiento posterior para que la representación en memoria se consolide y sea resistente al enmascaramiento (Potter, 1976). En un estudio más reciente, Greene y Oliva (2009) encontraron que el tiempo de exposición mínimo necesario para clasificar correctamente una imagen se situaba entre 19-67

ms, alcanzando un rendimiento asintótico a partir de los 100 ms. Otros estudios clásicos (e.g., Shepard, 1967; Standing, 1973) mostraron que no solo identificamos las imágenes rápidamente, sino que además podemos recordar un gran número de ellas a pesar de haberlas visto durante un breve instante.

Esa cantidad de información significativa que un observador puede extraer a partir de un vistazo y que le permite identificar y comprender lo que está viendo conforma lo que Oliva y Torralba (2006) denominan el *gist*<sup>1</sup> de la imagen. Así, el *gist* se debe entender como una representación espacial lo suficientemente completa como para captar el significado de la escena, reconocer algunos objetos y otra información saliente y facilitar la detección de objetos y el despliegue de la atención (Oliva, 2005).

El *gist* de una imagen incluye todos los niveles de información visual, por lo que Oliva (2005) establece una distinción entre *gist* perceptual y *gist* conceptual. El perceptual hace referencia a los niveles más básicos de la información visual, es decir, a las propiedades de la imagen que permiten realizar una representación estructural de las escenas (e.g. frecuencias espaciales, colores, tamaños, etc.). Retomando el ejemplo anterior, el *gist* perceptual contendría la información referente al tamaño de la cocina, su distribución espacial, o el color de los muebles. Este nivel de información visual estaría guiado por un procesamiento *bottom-up*, es decir, por la información perteneciente a los estímulos, y correspondería al procesamiento visual producido en los 100 ms iniciales de exposición.

Por otra parte, el *gist* conceptual incluye la información visual de alto nivel, como la activación del conocimiento semántico sobre la escena, la identificación de algunos objetos, o las posibles inferencias que el sujeto pueda realizar (Oliva, 2005). Así, en nuestro ejemplo, el *gist* conceptual sería la identificación de la cocina y la comprensión de su función, el reconocimiento del material con el que están fabricados los muebles, o la posible inferencia de que se trata de una cocina de estilo rústico. Este nivel se establecería en un momento temporal del procesamiento posterior a la formación del *gist* perceptual. Según Potter (1976), este intervalo temporal es necesario para que se produzca la consolidación de la imagen en la memoria, y está asociado a un procesamiento *top-down* basado, en gran medida, en las experiencias, conocimientos y expectativas de los observadores. Por ello, puede existir una

---

<sup>1</sup>Se ha optado por utilizar el término inglés debido a la dificultad para encontrar una palabra en castellano que represente fielmente este concepto. Algunas posibles traducciones son: “meollo”, quid, esencia, clave.

gran variabilidad entre los *gist* conceptuales que forman distintos espectadores que hayan estado expuestos a la misma imagen. Así, algunas personas pueden haber inferido que la cocina de nuestro ejemplo se encuentra en una vieja casa de campo rehabilitada, otras pueden haber creído que es una cocina de exposición en una tienda de mobiliario, y otro grupo puede haber pensado que se trata de la cocina de una casa abandonada.

## 1.1 Los componentes del *gist*

A la hora de estudiar qué tipos de información se incluyen en el *gist* de una imagen, en este trabajo adoptaremos la concepción de Wolfe, Horowitz y Michod (2007), quienes plantean que contiene, al menos, tres componentes diferenciados: 1) un conjunto de propiedades estadísticas; 2) información estructural sobre la composición de la imagen; y 3) un conjunto de objetos.

### 1.1.1 Propiedades estadísticas

Las propiedades estadísticas, características globales o estadísticas de totalidad, hacen referencia a la representación que el individuo elabora de los diferentes detalles de una imagen y a su combinación o asociación a través del espacio, incluso cuando estos detalles están contenidos en una frecuencia espacial determinada o cuando están unidos a objetos, partes, o localizaciones en el espacio. Dicha representación conforma un resumen a nivel estadístico de toda la información contenida en una escena (Álvarez y Oliva, 2009).

Nuestro sistema visual posee una gran capacidad para extraer estas propiedades estadísticas de las imágenes. Álvarez y Oliva (2008) mostraron que los sujetos son más precisos al determinar el centroide de un conjunto de elementos que al establecer la posición de un elemento individual. Chong y Treisman (2003) encontraron que los observadores eran capaces de extraer el radio medio de un conjunto de 12 círculos con una precisión similar a cuando juzgaban el radio de un único círculo; además, esta exactitud variaba mínimamente en función del tiempo de exposición (50-1000 ms) o de la demora de memoria (hasta 2s). Posteriormente, Chong y Treisman (2005a) probaron que esta precisión se mantenía independientemente del número de círculos que hubiese en el conjunto y de su densidad.

### 1.1.2 Información estructural

Oliva y Torralba (2001) entienden la imagen como un objeto individual antes que un conjunto de objetos. Así, estos autores acuñan el concepto de *spatial envelope* (“envolvente espacial”) para referirse a las propiedades espaciales existentes en una imagen que pueden ser extraídas a partir de computaciones simples, y que pueden traducirse en una descripción significativa del espacio de la escena. Según estos autores, el *spatial envelope* se define como: “el conjunto de límites, como paredes, secciones, suelos, elevaciones e inclinaciones de las superficies que definen la forma de un espacio”. Por otra parte, afirman que la información estructural de imágenes que pertenecen a la misma categoría semántica es similar, por lo que no sería necesario percibir los objetos en una escena para identificarla (Oliva y Torralba, 2001; 2006).

Las propiedades estructurales que definirían el *spatial envelope* de una escena y que, por tanto, serían suficientes para identificarla, serían: 1) el grado de naturalidad, 2) el grado de apertura, 3) el grado de rugosidad, 4) el grado de expansión y 5) el grado de robustez (Oliva y Torralba, 2001).

### 1.1.3 Objetos

Otro componente del *gist*, que nos facilita la identificación y el recuerdo de una imagen es la información que puede ser extraída de los objetos que contiene. Brady, Konkle, Álvarez y Oliva (2008) han encontrado que podemos recordar una gran cantidad de objetos presentados de forma aislada con un alto nivel de detalle, recordando, por ejemplo, su posición o estado (e.g., recordamos si una caja de cerillas estaba abierta o cerrada).

Sin embargo, parece que el recuerdo de imágenes no se basa en una lista de objetos aislados, pues dos imágenes considerablemente diferentes pueden dar lugar a una misma lista (Wolfe et al., 2007); además, se ha demostrado la relevancia del contexto en el reconocimiento de objetos en una escena (Hollingworth, 2006; Oliva y Torralba, 2007).

Por otra parte, se ha encontrado evidencia de que la información visual adquirida de los objetos atendidos durante la visualización de una escena se almacena en la memoria, independientemente de si el espectador tiene intención de recordarla o no (Castelhamo y Henderson, 2005). Más recientemente, Draschkow, Wolfe y Võ (2014) han mostrado que buscar objetos incluidos en escenas naturales (i.e., codificación incidental) crea representaciones de memoria más fuertes y duraderas que su memorización intencionada.

## 2. Atención visual y *gist*

A lo largo del día estamos constantemente percibiendo imágenes distintas que, normalmente, no necesitamos memorizar, pues la mayoría de las escenas que visualizamos son secundarias a la tarea que estemos realizando en cada momento. Por ejemplo, imaginemos que un familiar se ha mudado a una calle que no conocemos y nos ha invitado a ver su nueva casa. Cuando estemos de camino, puede que nos fijemos en los carteles numéricos encima de la puerta de cada casa, pero que no prestemos demasiada atención al color del edificio, su estilo arquitectónico, o el negocio que se encuentra en la planta baja. Imaginemos ahora que nuestra empresa organiza una cena de trabajo en un restaurante que no conocemos, y un compañero nos enseña una foto de su exterior; ¿seríamos capaces de recordar que hemos pasado por delante de ese establecimiento mientras buscábamos la nueva casa de nuestro familiar?

Los estudios que han tratado de explorar la importancia de la atención en el reconocimiento del *gist* de escenas naturales han demostrado que existe ceguera por inatención ante estos estímulos, es decir, los observadores son incapaces de percibir este estímulo cuando su atención está focalizada en otra tarea. Mack y Rock (1998) llevaron a cabo un experimento en el que se presentaba a los participantes una cruz durante un breve tiempo con enmascaramiento posterior y su tarea era indicar cuál de los brazos de la cruz era más largo. En el cuarto ensayo de cada bloque, la cruz se presentaba superpuesta a una imagen. La mayor parte de los sujetos fueron capaces de reconocer el *gist* de la escena, lo que los autores atribuyeron al tamaño de la imagen, pues este podría atraer su atención, que sería distribuida a través de todo el estímulo (Mack y Rock, 1998). A pesar de esta explicación, es posible interpretar estos resultados de forma alternativa, considerando la capacidad de extraer el *gist* de una imagen sin prestarle atención de manera expresa (Cohen, Álvarez y Nakayama, 2011).

Posteriormente, Cohen et al. (2011) afirmaron que los estudios previos que mostraban la posibilidad de percibir escenas naturales sin atención utilizaban tareas con escasas demandas cognitivas, dejando recursos atencionales disponibles para el procesamiento de la imagen. Así, utilizando dos tareas con una mayor demanda atencional (*multiple-object tracking* y presentación rápida de series visuales), y tras encontrar ceguera por inatención en el 57% de participantes, establecieron que, si la tarea requería grandes cantidades de recursos atencionales, este efecto se podía producir ante escenas naturales. Mack y Clarke (2012) obtuvieron resultados superiores a los de Cohen et al. (2011), encontrando ceguera por

inatención en el 83% de sujetos, lo que reforzó la hipótesis de que no es posible extraer el *gist* si no se presta atención a la imagen. Más recientemente, Ward y Scholl (2015) encontraron que la ceguera por inatención se produce incluso cuando los observadores informan en tiempo real de la aparición de un estímulo inesperado (sin oportunidad de un decaimiento de la información perceptual), por lo que establecen que la ceguera por inatención se debe a una limitación en la percepción, y no en la memoria.

En el estudio de los efectos de la disminución de la atención, uno de los paradigmas más utilizados es el de la tarea dual, que ha demostrado comprometer la memoria de imágenes (Intraub, Daniels, Horowitz y Wolfe, 2008). Rowe y Rogers (1975) encontraron que, cuando los participantes debían realizar una tarea de sombreado verbal de letras mientras veían una serie de imágenes, se producía un decremento en la memoria, tanto en tareas de recuerdo libre como de reconocimiento, en comparación con aquellos participantes que únicamente escuchaban las letras sin sombreado. Más recientemente, Wolfe et al. (2007) mostraron que los observadores manifestaban haber visto un menor número de imágenes cuando realizaban una tarea concurrente, produciéndose un mayor decremento cuando debían realizar una búsqueda visual en comparación con una tarea de detección auditiva.

## 2.1 La atención visual y los componentes del *gist*

En el apartado anterior se ha puesto de manifiesto la necesidad de prestar atención para poder elaborar y recordar el *gist* de una imagen, pero ¿de qué manera afecta la disminución de la atención al procesamiento de sus diferentes componentes?

En primer lugar, la evidencia experimental muestra que el recuerdo de las propiedades estadísticas de una imagen depende de la disponibilidad de unos recursos atencionales mínimos. Chong y Treisman (2005b) encontraron que el juicio del tamaño medio de un conjunto de estímulos era más preciso cuando se realizaba una tarea concurrente de atención distribuida, mientras que el juicio del tamaño de elementos individuales era más acertado cuando la atención se focalizaba en ese elemento. Álvarez y Oliva (2008) encontraron que, al determinar la posición particular de un estímulo no atendido, el rendimiento era azaroso, mientras que el juicio sobre el centroide de un conjunto de estímulos no atendidos era más exacto. Jackson-Nielsen, Cohen y Pitts (2017) determinaron que, a pesar de que algunas propiedades estadísticas como la diversidad de colores y tamaños o el tamaño medio de los

elementos puedan ser percibidas sin atención focal, se necesitan ciertos recursos atencionales para su percepción.

Por otra parte, se han encontrado resultados análogos en el caso de la información estructural. Oliva y Torralba (2006) consideran que la información estructural de una imagen puede ser tratada como una característica global, es decir, se crea una representación estadística de las propiedades espaciales de la imagen; así, al tratarse de una característica global, requeriría menos recursos atencionales. Posteriormente, se ha demostrado que, en condiciones de atención reducida, se perciben más los cambios que alteran la información estructural de la imagen que cambios de igual magnitud que no alteran su estructura (Álvarez y Oliva, 2009), lo que demuestra que este componente del *gist* puede ser representado en situaciones en las que los recursos atencionales son escasos.

En cuanto al recuerdo de objetos en estas condiciones de atención reducida, Wolfe et al. (2007) han encontrado que, cuando no forman parte de una estructura espacial significativa, los objetos son poco relevantes para el recuerdo de una escena.

### 3. El recuerdo del *gist*

Tal y como se ha expuesto, la extracción, comprensión y recuerdo del *gist* de una imagen se pueden producir incluso con exposiciones de tiempo muy breves, pero ¿cómo podemos explicar esta capacidad de los observadores para recordar una gran cantidad de imágenes tras haberlas visto durante solo unos milisegundos?

La memoria conceptual a corto plazo (*Conceptual Short Term Memory, CSTM*), propuesta por Potter (1976), puede darnos una posible explicación. Según esta autora, la CSTM se puede entender como un *buffer* mental en el cual los estímulos perceptuales y sus conceptos asociados en la memoria a largo plazo (MLP) son representados brevemente, permitiendo así identificar patrones, redes o estructuras significativas (Potter, 2012).

Así, teniendo en cuenta esta delimitación de la CSTM, cuando una persona percibe un estímulo significativo como una palabra o una imagen, este se identifica rápidamente (100-300 ms) y activa la información asociada contenida en la MLP. A esta información conceptual activada se añaden los enlaces que se crean entre los conceptos activos en la CSTM, así como el conocimiento de alto nivel que posee el observador y sus metas y expectativas, formando

una red que representa la comprensión de la información visual (Potter, 2012), como puede ser el *gist* de una imagen o el significado de una oración.

Sin embargo, no toda la información visual a la que el observador está expuesto da lugar a una red conceptual. La información conceptual, si bien se activa rápidamente, es altamente inestable durante las fases iniciales, y será desactivada en pocos cientos de milisegundos si no es incorporada a una red (Potter, 2012). Así, se produce un rápido decaimiento de la información que no permite establecer una red conceptual o que, a pesar de adherirse a una, no se selecciona para un procesamiento posterior. Por tanto, para que la información conceptual activada no decaiga, es necesario que se dé, al menos, una de las dos posibilidades siguientes: (1) que se utilice un proceso de atención selectiva dirigido por las propiedades conceptuales del input visual, o (2) que dicha información conforme una red conceptual o se adhiera a una existente (Potter, 2012). Ambas posibilidades ocurren en momentos muy tempranos de la cadena de procesamiento de la información visual, incluso antes de que la memoria operativa haya empezado a tratar esos estímulos. En el caso de que no se diera ninguno de ellos, la información conceptual no llegaría a introducirse en la memoria de trabajo, lo que llevaría a su desvanecimiento (Potter, 2012).

La CSTM está construida a partir de la experiencia y el aprendizaje, por lo que la comprensión inmediata es más probable que ocurra para el material familiar; es decir, nuestra habilidad para entender una imagen nueva en una fracción de segundo depende de nuestra experiencia visual previa (Potter, 2012).

En el caso de la comprensión de imágenes, los observadores extraerían el *gist* rápidamente, recuperando de la MLP información conceptual sobre la imagen y su contenido. Tras esta identificación, los observadores pueden incluir la representación de la imagen en la memoria operativa, por ejemplo, si tienen que describirla o identificarla en una tarea de reconocimiento; si no lo hacen, esta decaerá rápidamente. Si retomamos la distinción de Oliva (2005), la CSTM se encargaría del procesamiento intermedio que transformaría el *gist* perceptual en el *gist* conceptual.

Este trabajo tiene, precisamente, como objetivo general, explorar en qué medida el recuerdo de una imagen depende de los diferentes componentes de su *gist* en una situación en la que los recursos atencionales son escasos.

# Experimento 1

Con este experimento se trató de analizar en qué medida el recuerdo de una imagen no atendida depende de la información estructural. La restricción de la asignación de recursos atencionales a la imagen se llevó a cabo utilizando un paradigma de tarea dual habitual en investigaciones sobre memoria de imágenes en situaciones de disminución de atención (Intraub et al., 2008; Wolfe et al., 2007), tomando como tarea primaria una tarea de búsqueda visual serial, que demanda recursos atencionales y el uso de atención selectiva focalizada. La manipulación de la disponibilidad de información estructural en la imagen se llevó a cabo utilizando dos tipos de escenas: escenas completas, cuya información estructural estaba intacta, y escenas fragmentadas, cuya información estructural se comprometía al dividir la imagen en diferentes sectores y aleatorizar sus posiciones.

Teniendo en cuenta estas dos manipulaciones realizadas, el diseño de investigación utilizado es un diseño experimental factorial intrasujeto 2x2, con los factores: integridad de la imagen (completa, fragmentada) y tipo de tarea (tarea de memoria, tarea dual).

Las hipótesis que se pretenden poner a prueba con este experimento son las siguientes:

- 1) Existe una interferencia en el recuerdo de las escenas cuando los participantes deben realizar simultáneamente una tarea de búsqueda visual.
- 2) El recuerdo es mejor cuando las escenas mantienen intacta su información estructural (están completas).

## 1. Método

### 1.1 Participantes

La muestra utilizada estuvo formada por 18 participantes con edades comprendidas entre los 19 y 23 años ( $M=20.61$ ,  $DT=1.20$ ), de los cuales, 4 eran hombres ( $M=21$ ,  $DT=.82$ ) y 14 mujeres ( $M=20.5$ ,  $DT=1.29$ ), todos ellos alumnos de la Universidad de Santiago de Compostela. Todos presentaban visión normal o corregida con lentes graduadas o de contacto en el momento de la prueba y realizaron correctamente el *Test Ishihara* de ceguera al color. Su participación en el experimento fue voluntaria y desconocían el propósito de este.

### 1.2 Materiales y Procedimiento

Los estímulos consistieron en 280 imágenes divididas en dos grupos: escenas completas y escenas fragmentadas. En la Figura 1 se muestran ejemplos de ambos tipos de estímulos. El *gist* de las escenas completas se compone de información relativa a las propiedades estadísticas, información estructural e información de objetos, mientras que, en las escenas fragmentadas, existiría una ruptura en la información estructural, por lo que su *gist* se compondría únicamente de propiedades estadísticas e información de objetos.



**Figura 1.** Ejemplos de estímulos para el Experimento 1.

Todas las escenas fueron tomadas de la base de datos Places205 (Zhou, Lapedriza, Xiao, Torralba y Oliva, 2014). Los criterios utilizados para la selección de las escenas fueron:

que tuvieran una resolución mínima de 512x512 píxeles, que no hubiera grandes zonas amarillas o blancas y que no aparecieran personas en primer plano. Las versiones fragmentadas de los estímulos se consiguieron dividiendo cada imagen en 25 sectores mediante una matriz de 5x5 y aleatorizando la posición de cada uno de ellos. En el ejemplo de la Figura 2 se puede apreciar cómo se pierde gran parte de la información estructural, aunque es posible que se conserven ciertos objetos (e.g., un cartel, una puerta, una ventana). En todas las condiciones se superpuso sobre las imágenes una cuadrícula invisible de 5x5 sobre la que se colocaban, en posiciones aleatorias, los elementos de la matriz de búsqueda: ocho dígitos que podían ser, bien todos el número 2, o 1 o 2 de esos dígitos se sustituían por el 5; los números se presentaban en color amarillo y en formato digital (1.8°x0.9°). Así, en cada matriz de búsqueda, podía aparecer ningún estímulo objetivo, uno, o dos. Las imágenes comprendían un ángulo visual de 20°x20° y la excentricidad máxima posible de los elementos de la matriz de búsqueda fue de 14.14°.



**Figura 2.** Escena antes y después del fragmentado.

Los estímulos fueron presentados en un monitor ASUS VW195S de 19 pulgadas con una resolución de 1440x900 píxeles y una frecuencia de actualización de 60 Hz conectado a un ordenador Toshiba Satellite L70-B-10R. El experimento fue programado en Python, utilizando PsychoPy v3.0.0b12 (Peirce et al., 2019). Los participantes se situaron a una distancia de 57.4 cm del monitor.

El experimento comprendió cinco condiciones, cuatro de ellas resultado de combinar los dos tipos de estímulo con los dos tipos de tarea: tarea únicamente de memoria o tarea dual de memoria y búsqueda visual. La quinta condición consistió únicamente en una tarea de



En la Figura 3 se puede ver el desarrollo de los ensayos para ambas fases y ambos tipos de tarea. Durante la fase de entrenamiento, en las condiciones en las que la tarea era únicamente de memoria se pidió a los participantes que se concentrasen en mirar las imágenes que se les iban a presentar, mientras que en las condiciones de tarea dual se les pidió que realizaran la tarea de búsqueda, i.e., que dijese el número de cincos que aparecían sobre la imagen. Cada ensayo de esta fase consistió en la presentación, durante 500 ms, del estímulo con la matriz de búsqueda superpuesta, seguida de una pantalla de respuesta en la tarea dual ante la que los participantes debían indicar, mediante el teclado, cuántos cincos habían visto, o una pantalla vacía en la tarea únicamente de memoria. La distribución del tipo de ensayos en la tarea de búsqueda se hizo de la siguiente forma: no se presentaba ningún cinco en el 45% de los ensayos ( $n=18$ ), había un cinco en otro 45% de los ensayos ( $n=18$ ), y en el 10% de ensayos restantes ( $n=4$ ) se presentaban dos cincos. Los participantes, en ningún caso, conocían dichos porcentajes.

Durante la fase de test, se presentaba a los participantes una serie de imágenes y, utilizando el teclado, debían indicar para cada una de ellas si la habían visto previamente o si, por el contrario, consideraban que se trataba de una imagen nueva. El 50% de las imágenes mostradas en esta fase ( $n=20$ ) habían aparecido durante la fase de entrenamiento, mientras que el 50% restante ( $n=20$ ) eran nuevas.

Se ha aleatorizado el orden de aparición de las condiciones, así como el orden de presentación de los estímulos dentro de cada condición y la asignación de una imagen a la fase de entrenamiento o de test; de tal manera que, si una imagen era utilizada en la fase de entrenamiento o en la fase de test de una condición, no volvía a mostrarse en las condiciones restantes.

### 1.3 Análisis estadísticos

En primer lugar, se comprobó si la ejecución en la tarea de búsqueda visual de las condiciones de tarea dual era significativamente diferente al azar,<sup>2</sup> con el objetivo de constatar si los participantes siguieron las instrucciones y realmente habían focalizado su atención en la

---

<sup>2</sup> Se ha tomado como nivel de ejecución azarosa un 33.33% de aciertos. Basándonos en el Teorema de Bayes, calculamos la probabilidad de que un sujeto acierte en un ensayo sumando las probabilidades de que acierte en cada una de las tres respuestas posibles (0 cincos, 1 cinco, 2 cincos), teniendo en cuenta que el sujeto desconoce la proporción de respuestas correctas de cada tipo, de modo que  $p(A) = (.45*.33) + (.45*.33) + (.1*.33) = .33$ .

tarea de búsqueda y no en la imagen presentada de fondo. Esta comprobación se hizo mediante una prueba t para una muestra. Seguidamente, se aplicó una prueba t para muestras relacionadas con el fin de verificar si los participantes habían tenido un rendimiento similar en las tareas de búsqueda de las condiciones duales y de la línea base.

A la hora de analizar los datos de las pruebas de memoria se eliminaron aquellos ensayos en los que aparecían estímulos sobre los cuales la búsqueda visual fue errónea, es decir, aquellos ensayos en los que no se pudo confirmar que los sujetos estuviesen realizando la tarea primaria, por lo que pudieron haber dedicado recursos atencionales a la imagen. Para analizar estos datos se llevó a cabo una transformación de la VD, calculando el índice de sensibilidad  $d'$  a partir de la proporción de aciertos y falsas alarmas.<sup>3</sup> Para comprobar si existían diferencias significativas entre la ejecución en la tarea de memoria de las distintas condiciones se realizó un ANOVA 2x2, incluyendo los factores intrasujeto: integridad de la imagen (completa, fragmentada) y tipo de tarea (memoria, tarea dual).

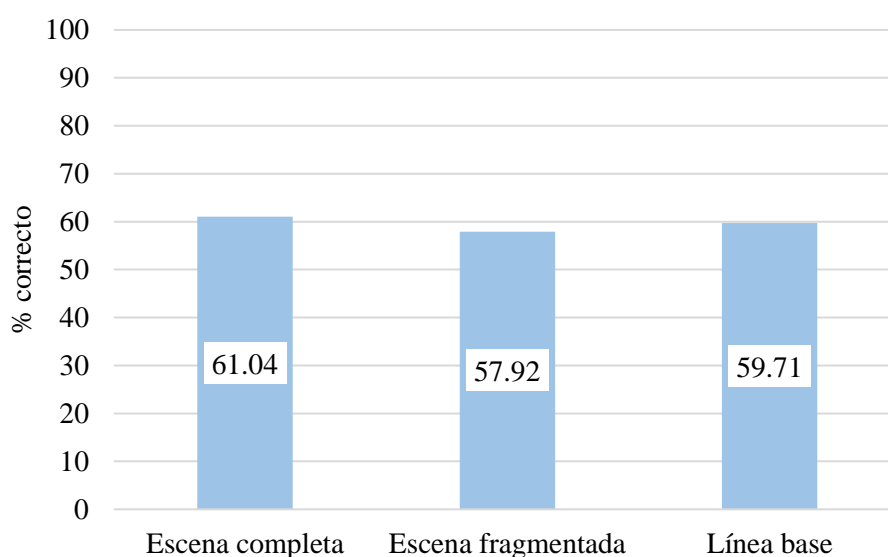
Todos los análisis se realizaron con el programa IBM SPSS Statistics versión 24 y se utilizó un nivel de significación de .05.

---

<sup>3</sup>  $d' = Z(\text{aciertos}) - Z(\text{falsas alarmas})$ . Este índice muestra la diferencia entre las distribuciones de aciertos y falsas alarmas, y permite determinar la exactitud con la que un individuo discrimina la señal del ruido.

## 2. Resultados

Tras hacer una primera revisión de los datos, fue necesario descartar los resultados de 6 de los 18 participantes: dos de ellos no dieron ninguna respuesta en más del 20% de los ensayos de, al menos, una de las condiciones; otros dos dieron una respuesta negativa en el 100% de los ensayos en, al menos, una de las condiciones; y los dos restantes emitieron respuestas azarosas (escaso número de aciertos y numerosas falsas alarmas). Por tanto, la muestra finalmente estuvo formada por 12 participantes, 3 hombres ( $M=21.33$ ,  $DT=.58$ ) y 9 mujeres ( $M=20.44$ ,  $DT=1.42$ ), cuyo rango de edad fue el mismo que el de la muestra inicial ( $M=20.67$ ,  $DT=1.30$ ).



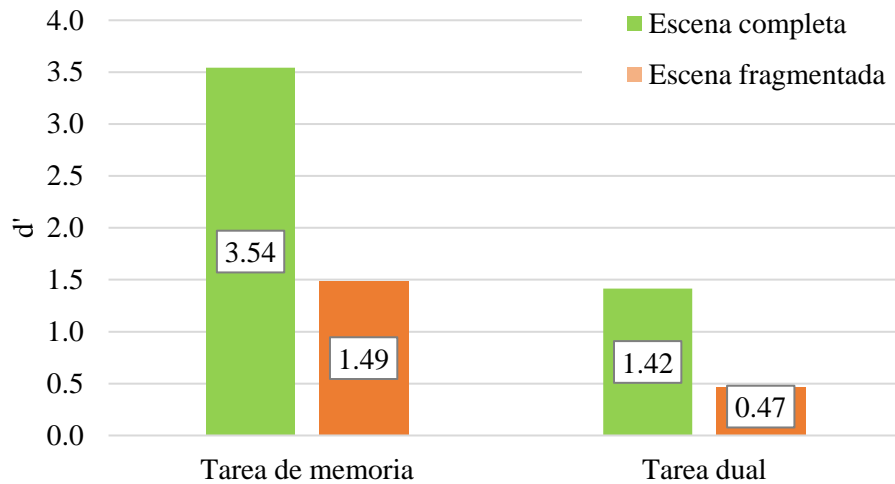
**Figura 3.** Ejecución media en la tarea de búsqueda visual del Experimento 1.

La Figura 4 muestra los porcentajes medios de aciertos en la tarea de búsqueda visual de las condiciones de tarea dual, así como en la condición de línea base. En los tres casos, se encontró una ejecución significativamente superior a la esperada por el azar (ver Tabla 1).

**Tabla 1.** Comparación entre la ejecución en la búsqueda visual y el azar en el Experimento 1.

	t	gl	Sig. (bilateral)
Escena completa	11.05	11	<.001
Escena fragmentada	14.88	11	<.001
Línea base	11.21	11	<.001

Por otra parte, no se han encontrado diferencias significativas en la ejecución en la tarea de búsqueda visual de las diferentes condiciones respecto a la línea base.



**Figura 4.** Rendimiento en la prueba de memoria del Experimento 1.

En cuanto a la ejecución en la tarea de memoria, en la Figura 5 se muestran los valores medios de  $d'$  en función del tipo de estímulo y del tipo de tarea. Se han encontrado efectos significativos en los factores: integridad de la imagen [ $F(1,11)=17.844, p \leq .001$ ] y tipo de tarea [ $F(1,11)=13.466, p \leq .004$ ], mientras que la interacción entre ambos factores no resultó significativa [ $F(1,11)=2.192, p \leq .167$ ].

### 3. Discusión

Los resultados obtenidos en el Experimento 1 nos permiten aceptar las hipótesis planteadas, ya que hemos encontrado que los participantes recuerdan un mayor número de escenas en la condición de tarea simple de memoria y, de igual manera, la cantidad de imágenes recordadas es superior cuando se presentaban íntegras.

En primer lugar, podemos observar cómo la ejecución en las tareas de búsqueda visual de las condiciones duales es similar entre condiciones, y es significativamente superior al azar, lo que nos indica que los sujetos estaban siguiendo las instrucciones y realizando la búsqueda. Estos resultados son semejantes a los obtenidos en otros estudios que utilizaron esta misma tarea de búsqueda visual (Intraub, 2008; Wolfe, 2007).

En cuanto a las pruebas de memoria, se aprecia una disminución significativa del recuerdo en las condiciones de tarea dual frente a las condiciones únicamente de memoria, es decir, los sujetos reconocían menos imágenes cuando debían realizar la búsqueda visual que, como se había previsto, les impedía atender a la imagen. Estos resultados son coherentes con los obtenidos por Wolfe et al. (2007), y reflejan la demanda atencional que se produce en una tarea dual y su efecto negativo sobre el recuerdo.

Por otra parte, se encontró que las escenas fragmentadas se recordaban significativamente peor que las completas en ambos tipos de tarea, lo que da cuenta de la relevancia de la información estructural en el recuerdo de imágenes, tanto en los ensayos en los que los recursos atencionales podían asignarse en su totalidad a la tarea de procesamiento de imágenes, como en aquellos casos en los que la cantidad de recursos disponibles estaba limitada como consecuencia de la presencia de la tarea de búsqueda visual, concurrente con la tarea de memoria.

De hecho, en las condiciones de tarea dual se puede observar cómo las escenas completas se recuerdan considerablemente mejor que las escenas fragmentadas aun cuando los recursos atencionales dedicados a la imagen son escasos. Estos resultados indican que la información estructural puede ser representada incluso en aquellas situaciones en las que los recursos atencionales son reducidos, tal y como se ha mostrado en estudios anteriores (Álvarez y Oliva, 2009; Oliva y Torralba, 2006).

## Experimento 2

En el Experimento 1 se mostró la relevancia de la información estructural a la hora de recordar una imagen cuando la atención estaba focalizada en otra tarea. A pesar de mostrar una peor ejecución cuando se corrompía este componente del *gist*, los participantes fueron capaces de identificar varias de las imágenes que habían visto, lo que nos indica que, aunque la información estructural esté comprometida, quizás los otros dos componentes del *gist* siguen proporcionando información suficiente para poder identificar una imagen.

En este experimento los participantes debían recordar texturas (imágenes cuyo *gist* no contiene información de objetos), con el objetivo de analizar en qué medida el recuerdo de una imagen no atendida depende de la información estructural en imágenes que no contienen objetos. La restricción de la asignación de recursos atencionales, así como la manipulación de la disponibilidad de información estructural en la imagen, se llevó a cabo del mismo modo que en el Experimento 1, por lo que el diseño utilizado fue el mismo.

Las hipótesis que se pretenden poner a prueba con este experimento son las siguientes:

- 1) Existe una interferencia en el recuerdo de las texturas cuando los participantes deben realizar simultáneamente una tarea de búsqueda visual.
- 2) El recuerdo es mejor cuando las texturas mantienen intacta su información estructural (están completas).

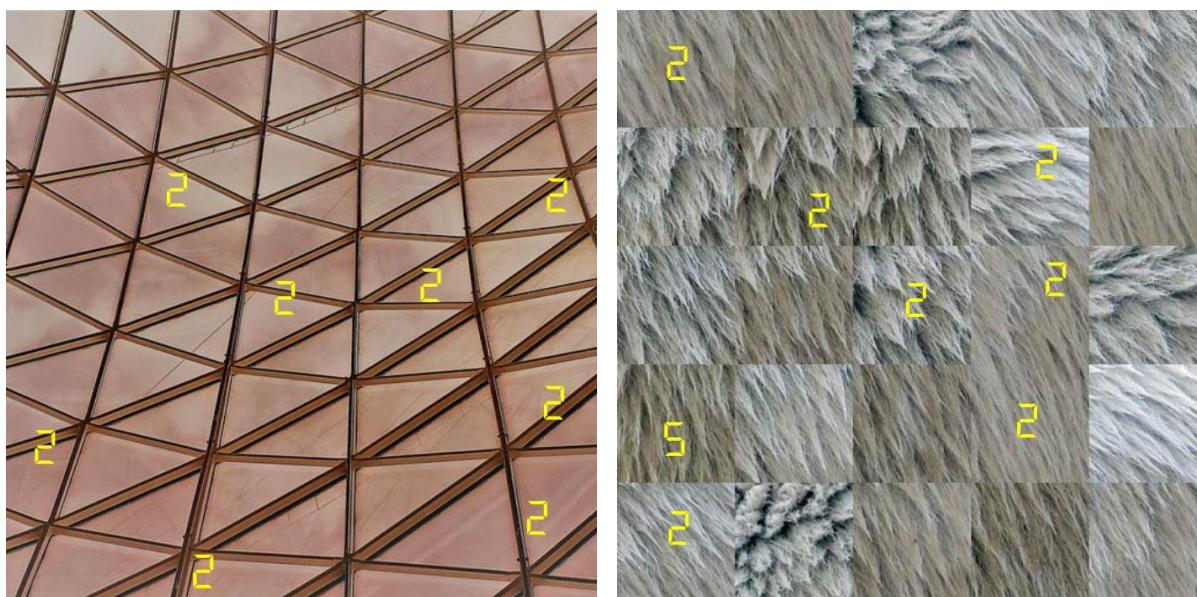
## 1. Método

### 1.1 Participantes

Los participantes de este experimento fueron los mismos que realizaron el experimento anterior.

### 1.2 Materiales y Procedimiento

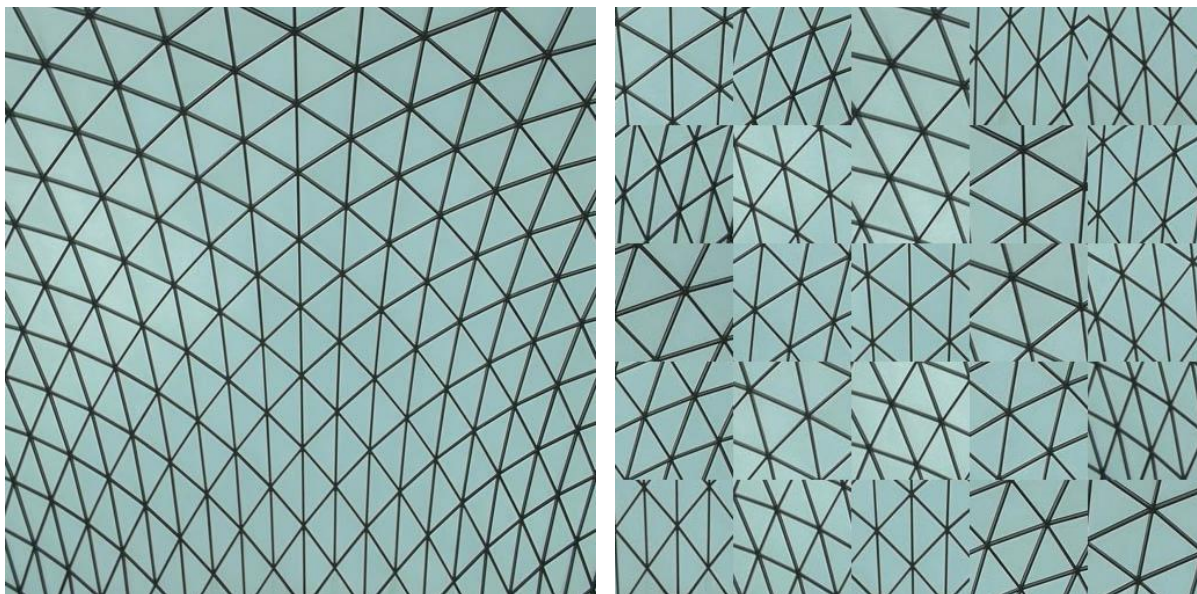
Para este segundo experimento se utilizó un total de 240 imágenes divididas en dos grupos de forma análoga al Experimento 1: texturas completas y texturas fragmentadas. A diferencia del experimento previo, las texturas completas se caracterizan por un *gist* constituido únicamente por propiedades estadísticas e información estructural, careciendo de objetos. Por otra parte, el *gist* de las texturas fragmentadas estaría compuesto tan sólo por propiedades estadísticas, ya que al ser una textura carece de objetos y al estar fragmentada se elimina, en gran medida, su información estructural. En la Figura 6 se muestran ejemplos de ambos tipos de estímulos.



**Figura 6.** Ejemplos de los dos tipos de estímulos usados en el Experimento 2.

Las texturas utilizadas fueron seleccionadas de la plataforma Pixabay usando el término de búsqueda “texture”. Los criterios para la selección de las imágenes fueron los mismos que en el Experimento 1, añadiendo la necesidad de que, en caso de existir dos texturas tomadas del mismo objeto (e.g. dos imágenes de corteza de árbol), estas fuesen fácilmente

diferenciables. La fragmentación de los estímulos se realizó del mismo modo que en el Experimento 1. En la Figura 7 se puede ver el resultado del proceso de fragmentación de una textura y cómo pierde la mayor parte de su información estructural. En todas las condiciones se superpuso una matriz de búsqueda sobre los estímulos con las mismas características que la usada en el Experimento 1. Se utilizaron los mismos aparatos que en el Experimento 1.



**Figura 7.** Textura antes y después del fragmentado.

En cuanto al procedimiento, en este segundo experimento se diseñaron cuatro condiciones que resultaban de cruzar los dos tipos de estímulo con las dos tareas: tarea de memoria o tarea dual. Como condición de línea base se utilizó la del experimento anterior. El procedimiento fue el mismo que el utilizado en el Experimento 1.

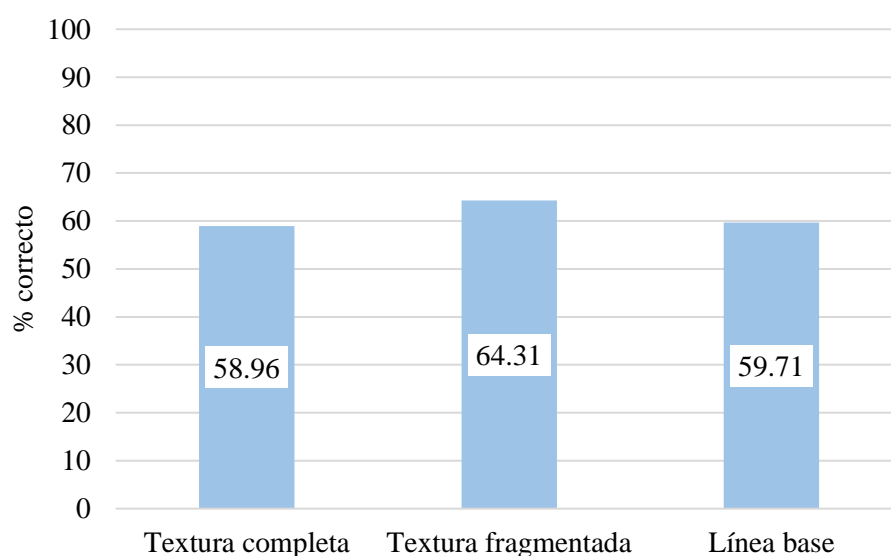
### 1.3 Análisis estadísticos

Los análisis estadísticos se realizaron de forma análoga al Experimento 1.

Con el objetivo de poder comparar los resultados de este experimento con los obtenidos en el experimento anterior, al realizar los análisis estadísticos no se tuvieron en cuenta los resultados de los seis participantes eliminados de la muestra del Experimento 1.

## 2. Resultados

La ejecución media de los participantes en la tarea de búsqueda visual de las condiciones de tarea dual se muestra en la Figura 8 en términos de porcentaje de aciertos. El rendimiento en las tres condiciones fue significativamente superior al azar (ver Tabla 2), y no se encontraron diferencias significativas entre la ejecución en las tareas de búsqueda visual de las condiciones de tarea dual y la línea base.

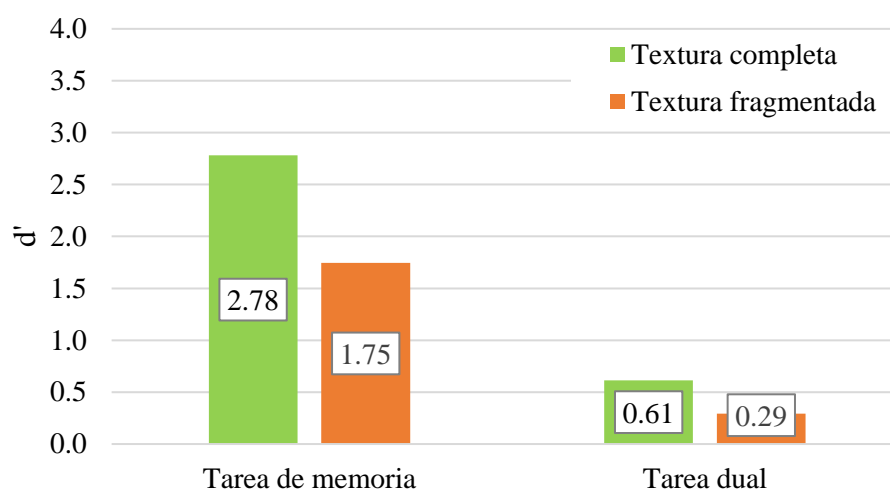


**Figura 8.** Ejecución media en la tarea de búsqueda visual del Experimento 2.

**Tabla 2.** Comparación entre la ejecución en la búsqueda visual y el azar en el Experimento 2.

	t	gl	Sig. (bilateral)
Textura completa	9.34	11	<.001
Textura fragmentada	14.61	11	<.001
Línea base	11.21	11	<.001

La Figura 9 recoge los resultados de los participantes en la prueba de memoria en función del tipo de estímulo y de la modalidad de la tarea. Se ha encontrado un efecto significativo del factor integridad de la imagen [ $F(1,11)=8.45$ ,  $p=.014$ ] y del factor tipo de tarea [ $F(1,11)=13.837$ ,  $p=.003$ ]; la interacción entre ambos no ha sido significativa [ $F(1,11)=3.879$ ,  $p=.075$ ].



**Figura 9.** Rendimiento en la prueba de memoria del Experimento 2.

### 3. Discusión

Los resultados mostrados en el apartado anterior nos permiten aceptar las hipótesis planteadas, puesto que se ha encontrado un mayor recuerdo de texturas en la condición de tarea simple de memoria y, de igual manera, se ha recordado un mayor número de imágenes cuando estas conservaban intacta su información estructural (estaban completas).

En primer lugar, la ejecución similar en la tarea de búsqueda visual de las distintas condiciones con respecto a la línea base, y superior al azar, nos indica que los sujetos estaban siguiendo las instrucciones y asignando recursos atencionales a esta tarea.

En cuanto a las pruebas de memoria, los resultados son análogos a los obtenidos en el Experimento 1. Así, observamos cómo el recuerdo de las texturas fue significativamente peor en las condiciones de tarea dual frente a las condiciones de tarea de memoria. Del mismo modo, los sujetos discriminaban peor las texturas fragmentadas que las completas, lo que nos indica que, aun cuando el *gist* de la imagen no contiene información sobre objetos, la información estructural es relevante para su recuerdo.

A pesar del bajo recuerdo de las texturas fragmentadas, se puede ver cómo los sujetos fueron capaces de reconocer algunas de ellas. En este caso, al fragmentar la imagen, y al no poseer objetos, ese recuerdo estaría basado únicamente en la información extraída de las propiedades estadísticas de la imagen, lo que nos mostraría que este componente puede ser percibido sin atención focalizada, tal y como plantearon Nielsen et al. (2017).

#### 3.1 Comparación entre experimentos

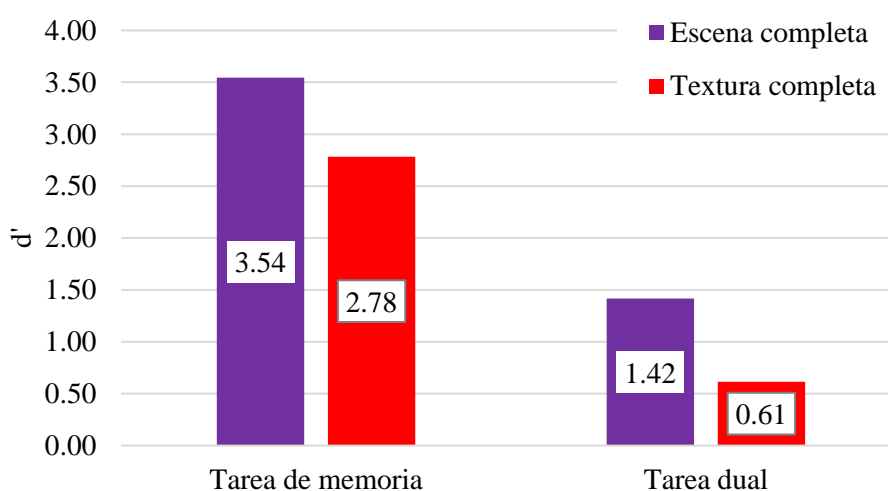
##### 3.1.1 Recuerdo de estímulos completos

Para explorar la importancia de la información extraída de los objetos en el recuerdo de imágenes en situaciones de atención limitada se realizaron comparaciones entre la ejecución en el primer experimento y en el segundo.

En primer lugar, se comparó la ejecución en la tarea de memoria de las condiciones de ambos experimentos en las que los estímulos estaban completos (i.e., conservaban su información estructural), explorando así las posibles diferencias en el recuerdo de imágenes cuyo *gist* poseía los tres componentes (escenas) con respecto a aquellas que no poseían

información de objetos (texturas). Los resultados de las tareas de memoria en esas condiciones se muestran en la Figura 10 en términos de  $d'$ .

Así, un ANOVA 2x2 sobre los factores estímulo (escena, textura) y tarea (tarea de memoria, tarea dual) arrojó un efecto significativo del factor tarea [ $F(1,11)=20.094$ ,  $p=.001$ ]. Estos resultados indican que los participantes recordaban significativamente menos imágenes cuando realizaban una tarea de búsqueda visual concurrente a la exposición, es decir, la demanda atencional de la tarea dual causó un decremento significativo en la memoria.

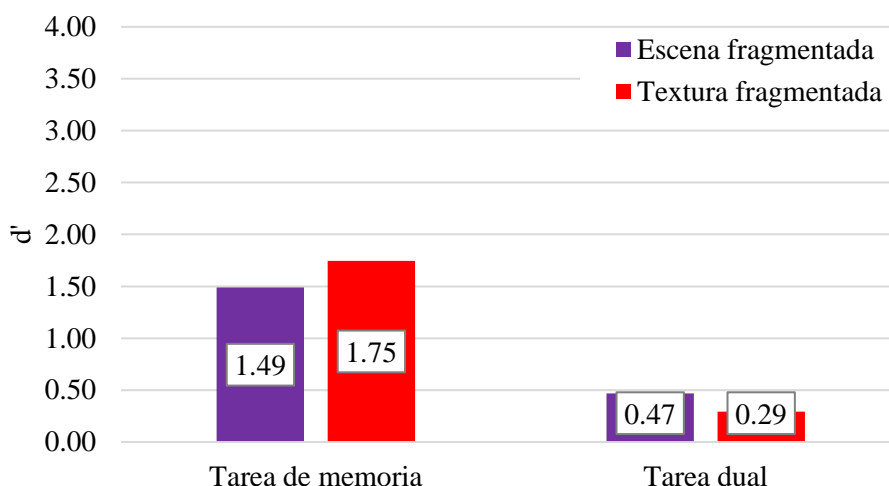


**Figura 10.** Rendimiento en función del tipo de estímulo y tarea en imágenes completas.

Por otra parte, se encontró un efecto significativo del factor estímulo [ $F(1,11)=6.182$ ,  $p=.03$ ], lo que sugiere que, cuando las imágenes presentaban una información estructural significativa, la información extraída de los objetos favorecía el recuerdo. No se ha encontrado una interacción significativa entre ambos factores [ $F(1,11)=.002$ ,  $p=.964$ ].

### 3.1.2 Recuerdo de estímulos fragmentados

Para analizar en qué medida el recuerdo de una imagen cuya información estructural está comprometida depende de la información extraída de los objetos, se llevó a cabo un ANOVA 2x2 con los factores estímulo (escena fragmentada, textura fragmentada) y tarea (tarea de memoria, tarea dual). Los valores medios de  $d'$  para estas condiciones se muestran en la Figura 11.



**Figura 11.** Rendimiento en función del tipo de estímulo y tarea en imágenes fragmentadas.

Los resultados muestran un efecto significativo del factor tarea [ $F(1,11)=13.768$ ,  $p=.003$ ], lo que nuevamente nos indica una peor ejecución en la tarea de memoria cuando los recursos atencionales estaban focalizados en la búsqueda visual. Sin embargo, el factor estímulo no ha resultado significativo [ $F(1, 11)=.027$ ,  $p=.872$ ], lo que pone de manifiesto que la información de objetos resulta irrelevante para el recuerdo cuando la imagen posee una información estructural dañada. Estos datos son semejantes a los obtenidos por Wolfe et al. (2007), y sugieren que el recuerdo de una escena no se limita a una lista de objetos, sino que estos deben adherirse a una estructura significativa para poder ser relevantes a la hora de recordar la imagen. La interacción entre ambos factores no ha resultado significativa [ $F(1,11)=.726$ ,  $p=.412$ ].

## Discusión general

Los dos experimentos descritos en este trabajo mostraron una clara disminución en el recuerdo de imágenes cuando estas se presentaban mientras el observador focalizaba su atención en una tarea de búsqueda visual. Estos resultados, típicos en un paradigma de tarea dual, muestran las limitaciones en la codificación de imágenes cuando los recursos atencionales disponibles son escasos (Wolfe et al., 2007). Dentro de estas condiciones de restricción atencional, los estímulos que se recordaron mejor fueron las escenas completas. Estas imágenes poseen los tres componentes del *gist*, y son las que contienen más información visual susceptible de ser recordada.

Por otra parte, los participantes recordaron de modo similar las escenas fragmentadas y las texturas completas, a pesar de que en las primeras no hubiera información estructural pero sí de objetos, al contrario que en las texturas. La CSTM (Potter, 1976) puede ayudar a interpretar estos resultados. Este modelo plantea que la información perceptual que se extrae de un estímulo establece nexos entre sí y con contenido relacionado alojado en la MLP, para crear una red conceptual significativa; en caso de que esto no ocurra, la información decae rápidamente. Es posible que, cuando los observadores ven escenas fragmentadas y texturas completas no sean capaces de establecer esos nexos con representaciones relacionadas en la MLP, lo que llevará a un decaimiento de estos estímulos.

En el caso de las escenas fragmentadas, puede ocurrir que, a pesar de poder conservar información relativa a objetos, los observadores no sean capaces de extraer la estructura de la

imagen y de llegar a su significado. Es decir, los sujetos perciben información conceptual (objetos) que, aunque sean capaces de asociar con representaciones individuales de dicho elemento en MLP, no puede llegar a crear una red conceptual, puesto que esa imagen probablemente no tenga asociada ninguna información en la MLP. Por otra parte, en las texturas podría ocurrir un proceso similar. Los observadores pueden extraer la estructura de la imagen pero, al no poder extraer ninguna información relativa a objetos, esa disposición estructural no sería suficiente para establecer un enlace con representaciones en la MLP, por lo que se produciría decaimiento.

Así, puede ocurrir que, en estos dos tipos de estímulos, al no ser posible el establecimiento de esa conexión con representaciones en la MLP y, consiguientemente, no poder dotar de significado a la información visual, el *gist* quede reducido al nivel perceptual y no llegue a elaborarse un *gist* conceptual.

Otro factor que puede haber influido en el bajo nivel de recuerdo de las imágenes que no presentaban los tres componentes del *gist* puede haber sido el breve tiempo de exposición. A partir de los 100 ms de exposición se produce un rendimiento asintótico en una tarea de clasificación de imágenes (Greene y Oliva, 2009), y a partir de los 300 ms posteriores la representación en memoria es resistente al enmascaramiento (Potter, 1976). En vista de estos datos, los tiempos de exposición utilizados en los experimentos aquí reportados (500 ms) deberían ser suficientes para la extracción del *gist*. Sin embargo, los resultados de Greene y Oliva (2009) y Potter (1976) se han obtenido utilizando escenas naturales, es decir, muestran los tiempos de exposición necesarios para el procesamiento adecuado de una imagen que contiene los tres componentes del *gist*. Por tanto, es posible que 500 ms resulten insuficientes para el procesamiento de imágenes cuyo *gist* no contiene información relativa a sus tres componentes, por lo que son más complejas. Esta posibilidad ha de ser explorada en futuras investigaciones.

Por último, es necesario considerar las limitaciones que presenta este trabajo. En primer lugar, cabe la posibilidad de que el supuesto teórico en el que nos basamos para establecer los componentes del *gist* de la imagen no sea el más adecuado. En línea con esto, es posible que la manipulación experimental llevada a cabo para modificar los componentes del *gist* no fuese tan efectiva como se buscaba. Cabe destacar también que, a pesar de haber estudiado el papel de dos de los componentes del *gist* (información estructural y objetos), no se ha analizado la relevancia de las propiedades estadísticas, por lo que la importancia de este componente en la

memoria de imágenes en situaciones de recursos atencionales limitados ha de ser explorada en futuros trabajos. Pese a ello, es probable que este componente se haya visto modificado al alterar la información estructural de la imagen. Finalmente, es necesario tener en cuenta el tamaño de la muestra, dado que, con un mayor número de participantes, los resultados podrían verse ligeramente modificados debido a la reducción de la variabilidad inter-sujeto.

## Conclusiones

En este trabajo se ha probado la limitación del sistema cognitivo humano a la hora de procesar y recordar información visual en situaciones en las que los recursos atencionales disponibles son escasos. Así, se ha mostrado que la codificación de imágenes en memoria es dependiente de las demandas cognitivas de otras tareas visuales simultáneas. Del mismo modo, ha sido patente la necesidad de los tres componentes del *gist* de una imagen para su recuerdo, mostrándose de forma más pronunciada en condiciones de atención limitada.

Por tanto, y teniendo en cuenta los resultados obtenidos en este estudio, podemos concluir que, para que una imagen pueda crear una red conceptual significativa, además de prestarle atención, debemos de poder extraer adecuadamente información de los tres componentes de su *gist*.

## Referencias bibliográficas

- Álvarez, G. A. y Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 19(4), 392-398. doi: 10.1111/j.1467-9280.2008.02098.x
- Álvarez, G. A. y Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7345-7350. doi: 10.1073/pnas.0808981106
- Brady, T. F., Konkle, T., Álvarez, G. A. y Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14325-14329. doi: 10.1073/pnas.0803390105
- Castelhano, M. S. y Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, 12(6), 1017-1040. doi: 10.1080/13506280444000634
- Chong, S. C. y Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393-404. doi: 10.1016/S0042-6989(02)00596-5
- Chong, S. C. y Treisman, A. (2005a). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, 45(7), 891-900. doi: 10.1016/j.visres.2004.10.004
- Chong, S. C. y Treisman, A. (2005b). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, 67(1), 1-13. doi: 10.3758/BF03195009

- Cohen, M. A., Álvarez, G. A. y Nakayama, K. (2011). Natural scene perception requires attention. *Psychological Science*, 22(9), 1165-1172. doi: 10.1177/0956797611419168
- Draschkow, D., Wolfe, J. M. y Võ, M. L. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, 14(8), 18. doi: 10.1167/14.8.10
- Greene, M. R. y Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, 20(4), 464-472. doi: 10.1111/j.1467-9280.2009.02316.x
- Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1), 58-69. doi: 10.1037/0278-7393.32.1.58
- Intraub, H., Daniels, K. K., Horowitz, T. S. y Wolfe, J. M. (2008). Looking at scenes while searching for numbers: Dividing attention multiples space. *Perception & Psychophysics*, 70(7), 1337-1349. Doi: 10.3758/PP.70.7.1337
- Jackson-Nielsen, M., Cohen, M. A. y Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition: An International Journal*, 48, 149-160. doi: 10.1016/j.concog.2016.11.007
- Mack, A. y Clarke, J. (2012). Gist perception requires attention. *Visual Cognition*, 20(3), 300-327. doi: 10.1080/13506285.2012.666578
- Mack, A. y Rock, I. (1998). *Inattention blindness*. Cambridge, MA: Bradford Books/MIT Press.
- Oliva, A. (2005). Gist of the scene. En Itti, L., Rees, G. y Tsotsos, J. K. (eds.), *Neurobiology of Attention* (pp. 251-256). Cambridge, MA: Academic Press.
- Oliva, A. y Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145-175. doi: 10.1023/A:1011139631724
- Oliva, A. y Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 155, 23-36. doi: 10.1016/S0079-6123(06)55002-2
- Oliva, A. y Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12), 520-527. doi: 10.1016/j.tics.2007.09.009

- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., [...] Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195-203. doi: 10.3758/s13428-018-01193-y
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 509-522. doi: 10.1037/0278-7393.2.5.509
- Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in psychology*, 3, 113. doi: 10.3389/fpsyg.2012.00113
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156-163. doi: 10.1016/S0022-5371(67)80067-7
- Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2), 207-222. doi: 10.1080/14640747308400340
- Ward, E. J. y Scholl, B. J. (2015). Inattention blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness. *Psychonomic Bulletin & Review*, 22(3), 722-727. doi: 10.3758/s13423-014-0745-8
- Wolfe, J. M., Horowitz, T. S. y Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research*, 47(7), 955-964. doi: 10.1016/j.visres.2006.11.025
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A. y Oliva, A. (2014). Learning deep features for scene recognition using places database. En Neural Information Processing Systems Foundation (Ed.), *Advances in neural information processing systems* (pp. 487-495). Lake Tahoe, NE: proceedings.

## Índice de figuras

<b>Figura 1.</b> Ejemplos de estímulos para el Experimento 1. ....	15
<b>Figura 2.</b> Escena antes y después del fragmentado.....	16
<b>Figura 3.</b> Ejemplo de ensayos para ambas fases y ambos tipos de tarea. ....	17
<b>Figura 4.</b> Ejecución media en la tarea de búsqueda visual del Experimento 1.....	20
<b>Figura 5.</b> Rendimiento en la prueba de memoria del Experimento 1. ....	21
<b>Figura 6.</b> Ejemplos de los dos tipos de estímulos usados en el Experimento 2.....	24
<b>Figura 7.</b> Textura antes y después del fragmentado.....	25
<b>Figura 8.</b> Ejecución media en la tarea de búsqueda visual del Experimento 2.....	26
<b>Figura 9.</b> Rendimiento en la prueba de memoria del Experimento 2. ....	27
<b>Figura 10.</b> Rendimiento en función del tipo de estímulo y tarea en imágenes completas.....	29
<b>Figura 11.</b> Rendimiento en función del tipo de estímulo y tarea en imágenes fragmentadas.	30

# Índice de tablas

**Tabla 1.** Comparación entre la ejecución en la búsqueda visual y el azar en el Experimento 1. ...20

**Tabla 2.** Comparación entre la ejecución en la búsqueda visual y el azar en el Experimento 2. ...26