Universidad Zaragoza
1542

Trabajo Fin de Máster

Estudio y aplicaciones de la percepción sensorial humana en Realidad Virtual

Study and applications of human sensory perception in Virtual Reality

Autor/es

Sandra Malpica Mallo

Director/es

Belén Masiá Corcoy
Ana Serrano Pacheu

Escuela de Ingeniería y Arquitectura

2018

TRABAJOS DE FIN DE GRADO / FIN DE MÁSTER

Escuela de Ingeniería y Arquitectura Universidad Zaragoza

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

D./Dª. Sandra Malpica Mallo,

con nº de DNI 72895299L en aplicación de lo dispuesto en el art. 14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster) Máster, (Título del Trabajo) Estudio y aplicaciones de la percepción sensorial humana en Realidad Virtual,

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 22 de Noviembre de 2018

Fdo: Sandra Malpica Mallo

Resumen

La realidad virtual (VR) es una nueva tecnología inmersiva que está llegando al nivel del consumidor hoy en día. La llegada de este nuevo medio conlleva la necesidad de generar nuevos contenidos. Para ello, es necesario entender las diferencias de VR con los medios tradicionales. Cuando el usuario se sumerge en esta tecnología, se encuentra con un entorno que cubre los 360 grados a su alrededor y en el que tiene una gran libertad de acción desconocida hasta el momento. Por ello, es de suma importancia estudiar cómo el usuario le presta atención al entorno virtual (VE), así como el estudio de técnicas de control de la atención. En esta interfaz entre el usuario y el VE se encuentra la interacción de varios sentidos. Una forma de poder controlar la interacción o la atención del usuario es mediante supresión perceptual. En concreto, este trabajo se centra en el estudio de la atención visual en VR, ya que es el sentido dominante con el que exploramos el VE. Durante este trabajo se mostrará cómo otro de los sentidos, el oído, es capaz de influir en la percepción visual en un entorno de VR hasta el punto de suprimirla durante unos instantes.

Existen distintos mecanismos de supresión visual bien conocidos. Ya sea debido al movimiento, a un *flash*, un parpadeo o un rápido movimiento del ojo, existen momentos en los que el ser humano no es consciente de los cambios en los estímulos visuales que ocurren a su alrededor. Estos mecanismos son utilizados en distintos ámbitos, pero normalmente requieren de un estímulo que los desencadene. En este trabajo se realiza un experimento con usuarios en el que se observa cómo seis tipos de sonidos distintos son capaces de causar una supresión visual por medio de efectos intermodales en una escena realista en VR. Este hecho da pie a sustanciales mejoras en las técnicas y metodologías existentes, como la eliminación de *hardware* añadido cuyo coste resulta prohibitivo para un usuario medio. La inspiración de este trabajo surge de la intersección entre varios campos: el entendimiento de los sistemas anatómicos y procesos fisiológicos que modelan cómo se procesa la información sensorial en el cerebro humano, junto con el conocimiento de los retos presentes en informática gráfica respecto a VR, han hecho posible que surja la idea de este trabajo.

Durante el experimento que se desarrolla en este trabajo, se demuestra que los seis tipos de sonidos probados (entre ellos una frecuencia pura, ruido blanco, marrón y rosa) pueden desencadenar la supresión visual, dispuestos en varias localizaciones fuera del campo visual del usuario. Durante esta supresión, los participantes no son capaces de *detectar* la aparición de un estímulo visual frente a ellos, ni de *reconocer* correctamente la forma del estímulo visual presentado en las pocas ocasiones en las que este es detectado. Una vez encontrado el efecto de supresión visual se analiza cómo diversos factores influyen en el mismo, y se observa una gran robustez del fenómeno de supresión en cuanto a todos los niveles de los factores comprobados. Se discute cómo se podría integrar este efecto en una nueva técnica de control de la atención en VR y cómo esta técnica podría ayudar a aumentar la eficacia del tratamiento de exposición para aracnofobia con un nuevo tratamiento que los pacientes podrían llevar a cabo por su cuenta, sin necesidad de contar con un terapeuta en todo momento.

Agradecimientos

Este proyecto se ha desarrollado en el *Graphics and Imaging Lab* de la Universidad de Zaragoza. Quisiera agradecer a todos sus miembros por seguir haciendo de nuestro laboratorio un lugar agradable en el que trabajar y compartir ideas. No imagino un lugar mejor que este en el que haber realizado mi TFM. Gracias a ellos y a todos los participantes del experimento por hacer este trabajo posible.

Me gustaría darles las gracias especialmente a mis directoras, Ana Serrano y Belén Masiá por su inquebrantable guía e incansable apoyo. Creo que cada vez soy más consciente del esfuerzo que supone dirigir un trabajo de estas características, y cada vez agradezco más todo el tiempo que han sacado para hacer de este trabajo lo que ha llegado a ser, pese a todas las obligaciones que tienen a la vez. También le estoy agradecida a Diego Gutiérrez, por darme la oportunidad de conocer este grupo y de quedarme aquí a realizar mi tesis doctoral, y sobre todo por su franqueza y buen trato.

Por otra parte, a nivel personal, me gustaría darles las gracias a todos mis amigos que un año más siguen prestándome su ayuda. A Estefanía Garijo por estar aún estando lejos, a Adrián Barranco por ayudarme a despejarme con escritos varios y noticias interesantes. A David, Álvaro y Sonia por seguir siendo *La Sección del Mal* (y sobre todo a esta última por corregir las primeras versiones de esta memoria). A todos los que nos juntamos para hacer barbacoas épicas en Montepinar, llueva o haga sol. A María Leoz, por seguir siendo la mejor hermana que podría desear y por no rendirse nunca.

Les doy las gracias a todos los integrantes de *la columna izquierda* por hacer este año tan ameno. Sin su ayuda no habría conseguido superar los baches del camino, y sin su compañía el viaje habría sido eterno. En especial gracias a Carmen Martínez por alumbrar hasta la hora más oscura con su alegría. A los profesores del máster por todo lo que me han enseñado y el esfuerzo que ponen en hacer las clases más entretenidas.

A mis padres por estar ahí día y noche apoyándome en todo lo humanamente posible y más allá. A David Leoz por seguir trabajando duro después de tantos años, llega la época en la que podré empezar a devolverte una parte de tanto esfuerzo. A Alfa, Beta y Baco, por la compañía y el amor incondicional.

Por último, quiero dedicarle este trabajo a mi abuelo Félix, que falleció en diciembre de 2017, y que siempre me animó a perseguir mis ambiciones académicas. Por enseñarme las cosas bonitas y sencillas de la vida, gracias.

Índice

Índice de figuras

1. Introducción

1.1. Contexto del proyecto

El auge de los sistemas de VR a nivel de consumidor ha permitido abrir nuevas líneas de investigación (ver Sección 2.2). Algunos trabajos se centran en paliar o eliminar las limitaciones y dificultades técnicas que siguen existiendo [1], otros en crear nuevas técnicas que mejoren la usabilidad de esta nueva tecnología [2]. Existe un tercer cuerpo de trabajos que abarca todas las técnicas o aplicaciones que no entran directamente en ninguno de los dos anteriores, si no que pueden trabajar de forma ortogonal a cualquiera de los dos para mejorar su aplicabilidad y potenciar su eficacia [3, 4]. Como en el caso de este trabajo, este último cuerpo busca una forma de aprovecharse de las peculiaridades de la interacción humana con el entorno.

La interacción humana con el entorno puede entenderse a partir de tres de sus sentidos: vista, oído y tacto. Estos tres sentidos forman parte de la interacción tanto con el mundo físico como con el entorno virtual (VE). En concreto, la vista y el oído son los que más estímulos del VE reciben. De forma análoga a la interacción con el entorno, definimos la atención visual que un humano presta a su entorno mayormente a partir de una información *a priori* simple: ¿Hacia dónde está mirando?

El conocimiento sobre la atención es valioso en sí mismo, porque nos aporta información sobre el comportamiento humano en distintos entornos. En medios tradicionales (y con el foco puesto en la atención visual), esta información ha resultado ser útil para campos tan diversos como marketing [5] (a la hora de colocar elementos atrayentes para el consumidor en los anuncios), generación de contenido [6], compresión de vídeo o imagen [7] (para preservar con mayor calidad las zonas en las que el usuario se va a fijar más), robótica [8], visión por computador [9] e incluso medicina [10].

Además, poder dirigir la atención visual del usuario permite ayudar al generador de contenido a comunicarse con el usuario de forma más eficaz, haciendo que le preste más atención a las zonas de mayor interés del medio que se le presenta [11]. También es posible aumentar sus capacidades discriminativas y de reconocimiento en tareas de distinta naturaleza, desde el entretenimiento hasta la medicina [12].

Ser capaz de modelar, predecir y controlar la atención del usuario cobra aún más importancia en el ámbito de la realidad virtual. De repente, el usuario adquiere un nuevo grado de libertad que no poseía en los medios más tradicionales. Al encontrarse en un entorno virtual inmersivo y que le rodea por completo, posee la opción de dirigir su atención a cualquier parte de ese

entorno, dejando de percibir la parte del mismo que queda fuera de su campo de visión (FOV). Hasta ahora, la forma física de presentación de los medios permitía en general contener toda la información del mismo dentro del FOV del usuario. Incluso al pensar en una pantalla de cine, un periódico o un escenario de teatro parece evidente que el usuario puede percibir toda la información necesaria de la manera en que ha sido diseñada por el generador de contenido.

Durante este trabajo se pretende conocer hasta qué punto los modelos, predictores y técnicas de control de la atención pensados para medios tradicionales son aplicables a VR. Se estudiarán sus limitaciones y se propondrán alternativas o nuevas técnicas para modular la atención en VR, discutiendo cuál o cuáles de dichas técnicas son aplicables en el ámbito médico actual.

Después de estudiar las técnicas de control de atención en medios tradicionales, queda claro que algunas de ellas se benefician de efectos de supresión perceptual (como las técnicas de *subtle gaze direction*). Por ello, se decide como un primer paso para la obtención de una técnica de control de la atención, encontrar un efecto de supresión perceptual consistente en VR. En concreto, debido a la fuerte influencia del sonido en los entornos inmersivos y a la necesidad de crear una técnica que no sea visualmente intrusiva, se busca que dicho efecto sea de carácter intermodal. Es decir, que un estímulo sonoro sea capaz de producir una supresión visual de algún tipo. Con este efecto conseguido, la propuesta de una técnica de control de la atención asociada es prácticamente inmediato.

1.2. Objetivo del proyecto

Este trabajo se sitúa en la intersección entre los campos de percepción e informática gráfica. Concretamente, se busca entender el funcionamiento de la percepción audiovisual y su relación con la atención en medios inmersivos y aplicar los conocimientos adquiridos a entornos de realidad virtual (VR). Su objetivo último es encontrar una idea o aplicación que permita mejorar procesos de rehabilitación motora o hacer más efectivo el tratamiento de exposición para fobias, ambas aplicaciones médicas que ya hacen uso de la VR [13, 14]. En concreto, se pretende demostrar que un estímulo auditivo es capaz de influir en la percepción visual hasta el punto de producir una supresión perceptual momentánea que hace que el sujeto no sea consciente de la información del entorno virtual en el que se encuentra inmerso por unos instantes. Poder inducir una supresión perceptual de esta forma implica la posibilidad de manipular el entorno virtual en el que se encuentra el usuario sin que éste sea consciente de ello, haciendo que su atención pueda ser controlada.

1.3. Alcance del proyecto

El alcance de este proyecto incluye:

- Recopilación del trabajo previo relacionado con la percepción audiovisual, modelos y predictores de la atención, técnicas de control de la atención y aplicaciones existentes relacionadas con VR.

- Estudio de la atención en entornos de VR. Análisis de técnicas existentes de control de la atención y sus límites.
- Propuesta de una nueva técnica de control de la atención basada en interacciones entre los sistemas visual y auditivo. En concreto, demostración de la existencia de un efecto de supresión visual causado por el sonido.
- Diseño, implementación y análisis mediante estudios de usuarios del efecto de supresión visual producido por una fuente de sonido para un conjunto controlado de condiciones.
- Discusión y planteamiento de su aplicación en procesos médicos de rehabilitación motora y tratamiento de exposición para fobias.

1.4. Organización del proyecto

Primero se explica el marco teórico básico necesario para comprender el resto del trabajo explicado a lo largo de la memoria en el Capítulo 2. Después, en el Capítulo 3 se presenta un breve resumen de los trabajos previos relacionados con modelado de la atención, predictores de la atención y técnicas de control de la atención en medios tradicionales además de distintos efectos de supresión perceptual. En el siguiente Capítulo (Capítulo 4) se presenta el experimento realizado en este trabajo y se discuten sus resultados (Capítulo 5). A continuación se relaciona el experimento con las técnicas de modelado y control de la atención en VR (Capítulo 6) y se propone cómo utilizar una técnica de control con el efecto demostrado a partir del experimento, aplicado al área de la medicina (Capítulo 7). Por último se exponen las conclusiones de este trabajo y se esboza la ruta a seguir como trabajo futuro.

1.5. Planificación

Durante los tres meses aproximados de duración del proyecto, el trabajo se ha dividido en las siguientes tareas: documentación del marco teórico, estudio del estado del arte, diseño, implementación y análisis de los dos experimentos realizados durante este trabajo, planteamiento de aplicaciones en el campo de la medicina, documentación y redacción de la memoria. La distribución de estas tareas a lo largo del tiempo se puede ver en la Figura 1.1.

Las tareas de implementación del experimento se han llevado a cabo principalmente en C# con el IDE Visual Studio asociado al programa Unity 3D. Tanto los estímulos visuales como auditivos han sido obtenidos con licencias Copyleft, que permiten su uso libre y gratuito. El análisis de los resultados se ha llevado a cabo con Matlab y R. Se han realizado copias de seguridad tanto para la escena completa del experimento como para el código implementado, junto con un sistema de versionado de archivos en Google Drive.

En cuanto al diseño del experimento, se ha seguido un método iterativo en el que en cada iteración se presentaba y discutía un borrador del experimento y se proponían mejoras y nuevas soluciones para la siguiente iteración. Finalmente, con la propuesta final de diseño del experimento se realizó un piloto con dos participantes para ajustar los distintos parámetros asociados

al mismo y mantenerlos fijos para el resto de participantes, así como para comprobar si se producía el efecto esperado y para asegurar que no hubiese fallos de implementación durante el experimento.

Etapa	Proceso	Horas
Trabajo previo	Documentación del marco teórico	25
	Estudio del estado del arte	37
Atención en VR	Análisis de técnicas existentes	37
	Estudio de limitaciones y mejoras	13
Experimento	Diseño	75
	Implementación	80
	Desarrollo	23
	Análisis	43
Propuesta	Nueva técnica de control	12
	Tratamiento de exposición	8
Documentación y bibliografía	Bibliografía	11
	Redacción de la memoria	37
		401

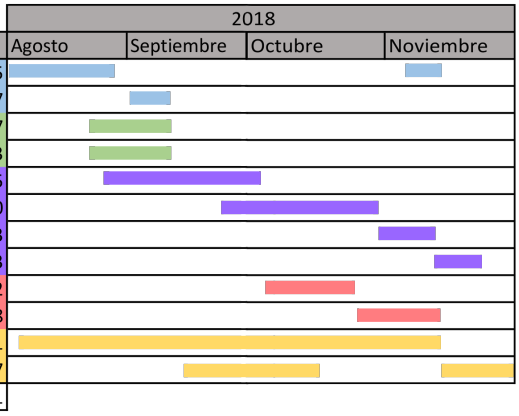


Figura 1.1: Diagrama de Gantt de las actividades realizadas a lo largo del proyecto.

2. Marco teórico

En este capítulo se describen varios conceptos, tanto teóricos como prácticos, que se utilizan a lo largo de este trabajo. La intención de este capítulo es la de ayudar a hacer de este documento un conjunto de información autocontenida. Sin embargo explicar en profundidad todos los conceptos que aquí se utilizan sería demasiado extenso. Por ello, se explican los fundamentos de cada concepto que se consideran suficientes y necesarios para sustentar la teoría utilizada en este trabajo. Para una explicación más detallada se recomienda consultar la bibliografía de este trabajo.

2.1. Los sistemas sensoriales

Los sistemas sensoriales (externos) son los encargados de procesar distintos tipos de energía para convertirla en información relevante acerca de parte de la realidad circundante a un individuo. Se reciben distintos tipos de estímulos del exterior que son procesados y desencadenan diversas respuestas, visibles o no, tanto en el cuerpo del receptor como hacia su entorno. En la Figura 2.1 se puede observar de forma resumida una distribución de las zonas de procesado final del *input* sensorial de los dos sentidos o modalidades en los que este trabajo se centra: la auditiva y la visual. Cada sistema sensorial se apoya en órganos o células altamente especializadas para transformar la energía de los estímulos en pulsos eléctricos capaces de viajar por el sistema nervioso.

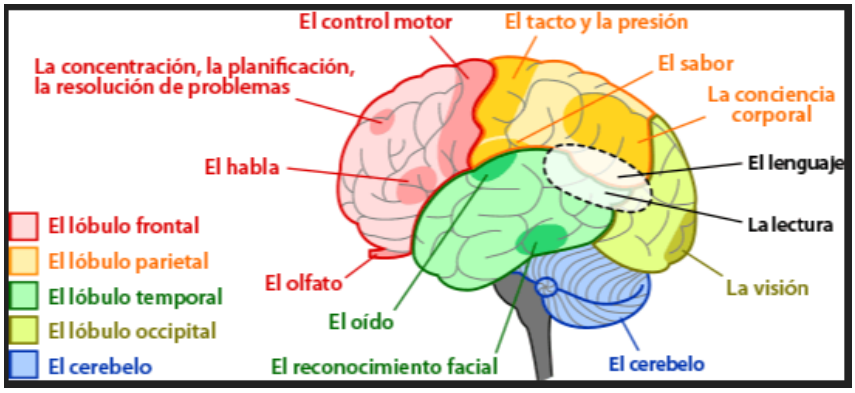


Figura 2.1: Esquema de las zonas de la corteza cerebral en las que se terminan de procesar los estímulos sensoriales externos que recibe el cerebro humano. También aparecen algunas zonas de respuesta asociadas a estos estímulos, como la corteza motora o el área del habla. Imagen de [15].

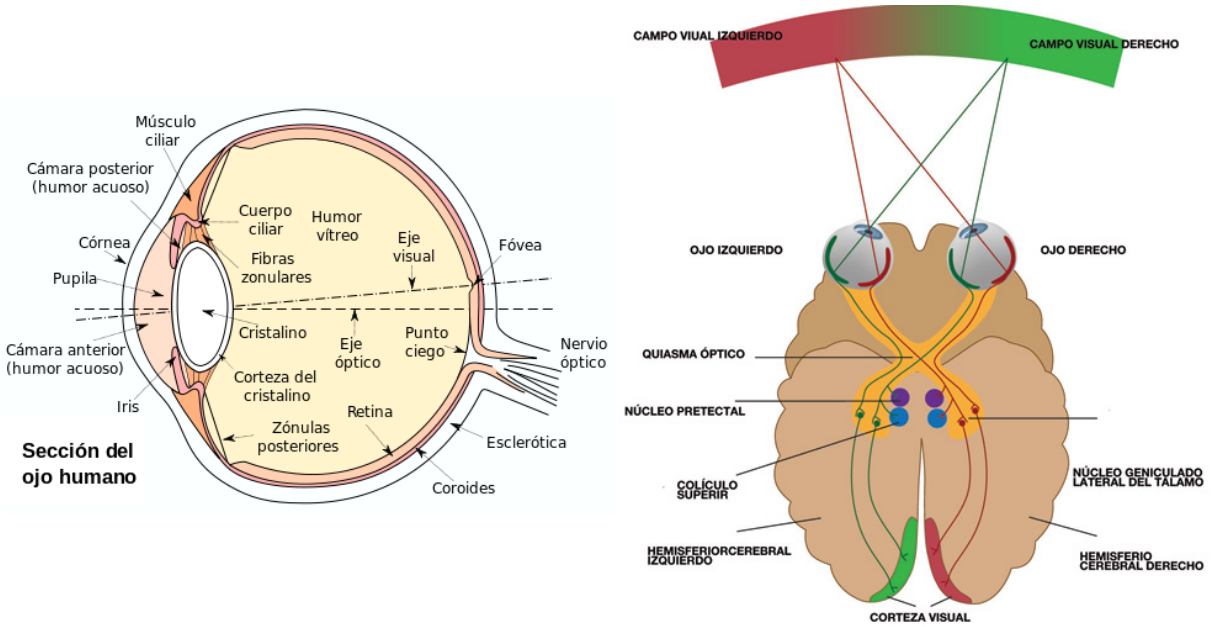


Figura 2.2: *Izquierda:* Esquema anatómico del ojo. *Derecha:* Camino que sigue la información relacionada con los estímulos visuales hasta llegar a la corteza cerebral. Imagen de [16].

Sistema visual. En el caso del sentido de la vista, su órgano especializado es el ojo. En la Figura 2.2 se puede ver un esquema de este órgano y del camino que siguen los impulsos producidos por los estímulos visuales hasta alcanzar la corteza del lóbulo occipital.

Los rayos de luz penetran en el ojo a través de la córnea y a continuación pasan a través de la pupila. La luz continúa su camino en un nuevo medio (el humor vítreo) hasta que llega a la retina. El punto de mayor concentración de rayos de luz en la retina es la fóvea, el punto de máxima visión. A lo largo de la retina (y con mayor concentración en dicho punto) se encuentran los conos y los bastones, que son células áltamente especializadas capaces de transformar la información luminosa en impulsos eléctricos que llegan hasta el nervio óptico y salen del ojo hacia el cerebro. El nervio óptico penetra en el cráneo y se dirige hacia el quiasma óptico (situado en la base del encéfalo), lugar en el que se cruzan los dos nervios ópticos. En el quiasma algunas fibras del nervio óptico cambian de hemisferio según la zona de la retina en la que se ha recibido el estímulo, como se puede ver en la Figura 2.2. Durante este recorrido de la información se produce un primer procesado en el núcleo geniculado lateral del tálamo, y a partir de ese punto surgen las radiaciones ópticas que viajan hasta la corteza visual donde se obtiene finalmente la información del entorno. Gracias a las diferencias percibidas por cada uno de los ojos se puede estimar la profundidad a la que se encuentran los objetos vistos.

En cuanto a la forma en la que los ojos se mueven, podemos distinguir entre dos tipos de patrones: por un lado, existen las sacadas, que son movimientos rápidos del ojo entre dos puntos. Por otro lado, se producen las fijaciones, que tienen lugar cuando el ojo se encuentra *inmóvil* o fijo en un objeto o zona a la que se está prestando atención visual.

Sistema auditivo. En el caso del sonido, el órgano que procesa los estímulos hasta transformarlos en señales eléctricas es el oído. El oído se compone de tres partes: oído externo, medio e interno, como se puede observar en la Figura 2.3.

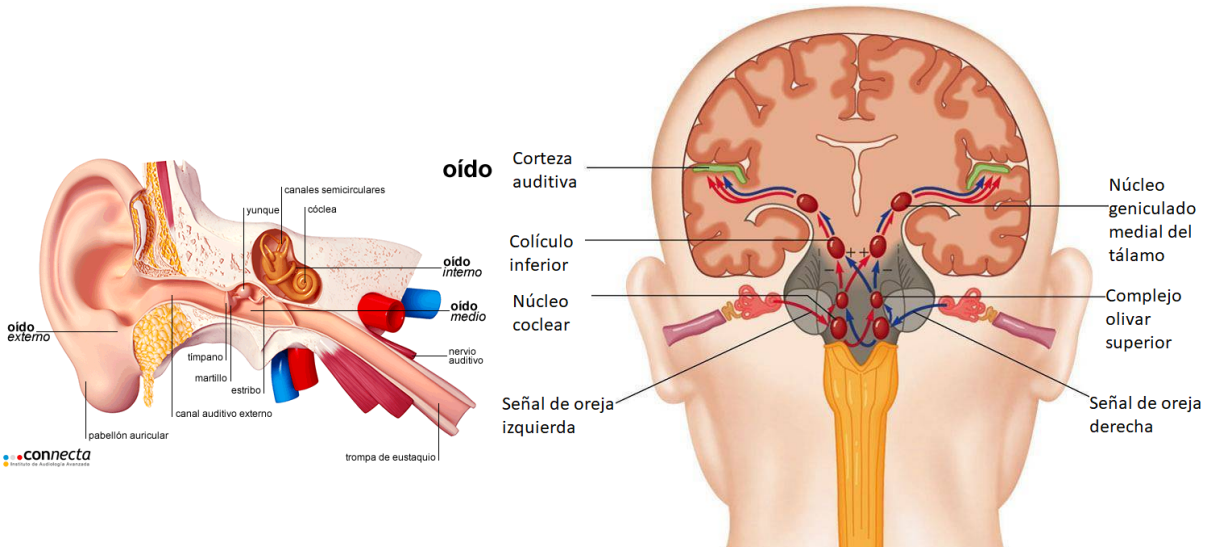


Figura 2.3: *Izquierda:* Esquema anatómico oído. *Derecha:* Camino que sigue la información relacionada con los estímulos sonoros hasta llegar a la corteza cerebral. Imagen de [17].

La oreja actúa como un receptor de vibraciones (ondas sonoras) que captura y transmite a través del conducto auditivo hasta el tímpano. Al llegar al tímpano, la vibración se transfiere hacia los huesos del oído medio, que la amplifican y conducen hacia la ventana oval del oído interno. En este punto se produce un cambio al medio líquido de la misma forma que en el sistema visual la luz llega al humor vítreo tras atravesar la pupila. En el oído interno este líquido (perilinfa) estimula las células cilíacas dentro de la cóclea. Estas células son las encargadas de transformar el estímulo recibido en impulsos eléctricos que viajarán a través del nervio auditivo hacia la corteza auditiva, pasando por el núcleo geniculado medial del tálamo. La diferencia temporal de las vibraciones que llegan a cada uno de los oídos permite localizar la dirección de la fuente que los ha originado.

¿Cómo se relacionan ambos sentidos entre sí? La información sonora y visual se puede complementar para ofrecer una mejor idea del mundo que nos rodea. Los estímulos procesados por sus sistemas asociados se mezclan o coordinan para estimar con mayor precisión el movimiento de un objeto [18], para poder cruzar con mayor seguridad una calle abarrotada o para entender mejor lo que dice una persona [19]. A pesar de que el sentido visual humano parece ser dominante sobre el resto debido a sus ventajas a la hora de obtener información espacial del entorno [20], el sentido auditivo muestra un valor único al suplir las carencias del primero, por ejemplo informándonos de lo que ocurre fuera de nuestro campo de visión, en la oscuridad o detrás de cualquier objeto que bloquee nuestra visión [21]. A nivel neurológico, se ha demostrado que existen zonas del sistema nervioso que procesan ambos tipos de estímulos o que se ven afectados por las dos modalidades [22, 23] y que nuestra percepción puede verse alterada cuando se procesan estímulos provenientes de distintas modalidades que muestran incongruencias [24]. Conocer la existencia de estas interacciones multimodales permite sacar el máximo partido a las particularidades de nuestros sistemas perceptuales y crear aplicaciones más efectivas o sorprendentes para el usuario.

Figura 2.4: Sistema de VR completo de HTC Vive. Consta de unas gafas de VR (centro) que se colocan en la cabeza y contienen tanto una pantalla para cada ojo como unos auriculares que transmiten el sonido de forma binaural, dos mandos (inferior izquierda y derecha) con los que el usuario puede interactuar con su entorno y dos *trackers* (superior izquierda y derecha) o localizadores que sirven para realizar el posicionamiento y seguimiento del jugador en la estancia física. El movimiento capturado puede ser trasladado al VE de tal forma que el usuario se vea reflejado en el mismo de una forma más realista. Imagen obtenida de la página oficial de HTC Vive.

2.2. La realidad virtual

La realidad virtual (VR) es una tecnología multimedia surgida en la década de los 90. Sin embargo, debido a limitaciones tanto de *hardware* como de *software* su llegada al consumidor se ha visto retrasada hasta hace pocos años. El auge de la VR viene acompañado por una explosión en nuevas aplicaciones que se aprovechan de esta tecnología (en educación [25], medicina [26], entretenimiento [27], etc.) y nuevas líneas de investigación que buscan aumentar la frontera del conocimiento gracias a su uso. La VR constituye un medio nuevo y fundamentalmente diferente a los medios convencionales [28] que se caracteriza por un alto grado de inmersión y presencia, además de nuevas formas de interacción con el entorno virtual (VE).

Existen muchos distribuidores de gafas de VR, pero uno de los más conocidos y el utilizado en este trabajo es el sistema de realidad virtual de HTC Vive (Figura 2.4). Las especificaciones de este sistema de VR se pueden ver en la Tabla 2.1. El *head mounted display* (HMD) presenta integrado un *eyetracker* de la empresa Pupil Labs (Figura 2.5), cuyas especificaciones se pueden consultar en la Tabla 2.2. Los *eyetrackers* son aparatos utilizados para medir la posición de la pupila y el movimiento ocular. Se utilizarán en este trabajo para obtener datos adicionales de los participantes del experimento.

Los sistemas de VR permiten a los usuarios sentirse inmersos en los VE sin requerir un alto grado de concentración. Cuando el usuario se encuentra en VR se cumple el principio de suspensión de la incredulidad que le permite disfrutar de la experiencia dejando a un lado su sentido crítico. Los *trackers* del sistema permiten una interacción más natural con el entorno,

Pantalla	AMOLED dual 3.5”
Resolución	1440x1600 píxeles por ojo (2880x1600)
Tasa de refresco	90Hz
Campo de visión (FOV)	110 grados

Tabla 2.1: Especificaciones técnicas del sistema de VR HTC Vive.

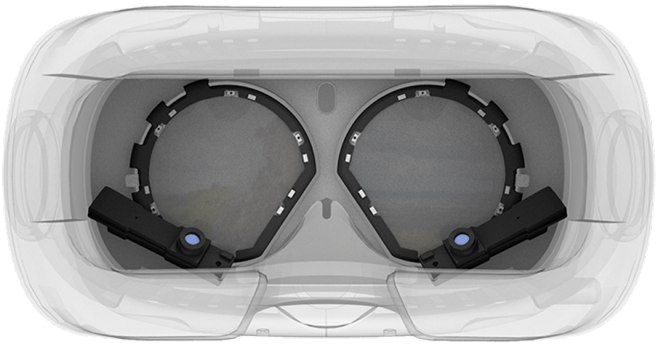


Figura 2.5: Montaje del *eyetracker* de Pupil Labs en el sistema de VR de HTC Vive. El *eyetracker* consiste en dos cámaras de visión infrarroja montadas cada una en un anillo con varios LEDs que se colocan sobre las lentes de las gafas de VR y apuntan directamente al ojo del usuario. Imagen obtenida de la web oficial de Pupil Labs.

lo que aumenta el grado de inmersión. Pese a las grandes ventajas de este nuevo medio se debe tener en cuenta que cualquier desajuste entre lo que el usuario espera percibir debido al VE y lo que percibe realmente producirá un malestar físico en el usuario (*simulator sickness*).

Las diferencias entre VR y el resto de medios convencionales son también las razones de que surjan grandes retos de investigación a su alrededor. Una forma de mejorar la experiencia del usuario y evitar los problemas asociados a esta tecnología es utilizar un mejor *hardware*. Otra, es tener en cuenta las particularidades del sistema sensorial humano y aprovecharlas para crear aplicaciones de forma más inteligente y eficaz. Los trabajos de grandes grupos de investigación de reconocida fama internacional como Nvidia e Intel [3, 4] además del grupo *Graphics and Imaging Lab* de la Universidad de Zaragoza, del que la autora forma parte, sugieren que este es un camino prometedor y potencialmente beneficioso para el campo [29, 30].

Frecuencia de muestreo	120Hz
Campo de visión (FOV)	Hasta los límites de HTC Vive
Precisión	0.08 grados
Latencia de cámara	5.7ms
Resolución	640x480
Método de calibración	2D - 7 puntos

Tabla 2.2: Especificaciones técnicas del sistema de *eyetracker* de Pupil Labs montado en el HMD de HTC Vive.

3. Trabajo relacionado

A continuación se presenta un resumen del trabajo relacionado con este proyecto. Para facilitar la comprensión del mismo, éste se presenta agrupado por distintas categorías, cada una de ellas relacionada con un aspecto concreto o la totalidad del trabajo que se desarrolla en este documento. Esta sección no pretende ser exhaustiva, si no dar una idea general de algunos campos relacionados con este trabajo.

Supresión perceptual e interacciones intermodales. Existen diversos trabajos que se dedican a estudiar cómo funcionan los mecanismos responsables de la supresión perceptual. Existen fenómenos bien conocidos, como la rivalidad binocular [31], la ceguera inducida por el movimiento [32], la supresión causada por un *flash* [33] o supresión sacádica [34]. Existe también la supresión perceptual intermodal, como la ilusión de la linterna [24] en la que un sonido influye en la percepción de continuidad de un estímulo visual. Estos fenómenos sirven, entre otras cosas, para estudiar las diferencias entre el procesamiento consciente e inconsciente de los estímulos visuales, o para utilizar estímulos visualmente *invisibles* en un experimento. En el caso del control de la atención, la supresión visual permite modificar el entorno del usuario sin que este sea consciente de los cambios. Se ha demostrado recientemente que las interacciones intermodales se mantienen en VR [35]. De hecho, existen varios experimentos que utilizan la supresión perceptual para dirigir, controlar o predecir la atención del usuario en VR como los trabajos de Arabadzhiyska et al. [4], Sun et al. [3], Langbehn et al. [36] y Bolte et al. [37]. Estos trabajos se explican más a fondo en la Sección 3.1.

Modelos de atención convencionales. Existen muchas formas de modelar la atención humana. Por un lado hay modelos analíticos que tratan de predecir la atención [38]. Por otro, colecciones de descubrimientos o reglas que describen con rasgos generales el comportamiento humano. Algunas se centran en el oído [39] pero la mayoría de ellas se centra en la atención visual [40], ya que la vista es el sentido dominante en los humanos (Sección 2.1). Otros trabajos tienen en cuenta más de una modalidad, como la audiovisual [41]. Una de las teorías aceptadas es la de Wolfe y Gray [42], que clasifica las características que más llaman la atención de los estímulos visuales en tareas de búsqueda. El trabajo de Ditterich et al. demuestra que la atención visual y las sacadas están correladas [43, 44]. La atención visual también se relaciona con el sonido [45].

Una forma de predecir dónde se fijará alguien es calcular un mapa de saliencia. La saliencia es una característica propia de un objeto que indica cuánto resalta comparado con su entorno local. Los mapas de saliencia pueden calcularse a partir de las propias características de la imagen, teniendo en cuenta por ejemplo los factores que nos llaman la atención de la teoría de Wolfe [46]. Otra forma de calcularlos es la que se utiliza en el trabajo de Sitzmann et al. [29], donde el mapa de saliencia de una escena se calcula promediando las observaciones reales de un conjunto

de personas. Otra forma de estudiar en qué partes y en qué orden se ha fijado alguien son los *scanpaths*. Estas visualizaciones pueden aportar información extra sobre la duración de cada fijación y el patrón espacial seguido [47].

Podemos entender por control de la atención una modificación intencionada de la misma, ya sea haciendo que el objetivo de la atención cambie o que pase por alto ciertos eventos. Un parpadeo de alta frecuencia presentado en una imagen puede captar nuestra atención [48]. Si alguien grita nuestro nombre en medio de una fiesta nos llamará la atención (*cocktail party effect* [49]). Todas estas técnicas pueden clasificarse como sutiles o invasivas. Las primeras ocurren sin que el usuario sea consciente de que ha recibido un estímulo que le ha hecho modificar su atención [50]. Las segundas utilizan estímulos obvios que llaman la atención de forma abierta [51]. Las técnicas sutiles incluyen seguir la mirada de otra persona [52], ligeras modulaciones de color o luminancia [53], parpadeos de una frecuencia determinada [54] o aplicación de filtros de *blur* sobre la escena [11]. Las técnicas invasivas incluyen resaltar con cajas rojas los objetos a los que hay que prestarles atencion o modificar su tamaño [55]. La atención auditiva se puede mejorar con estímulos visuales [56] y viceversa [57].

Uso de VR en medicina. La realidad virtual ha demostrado ser un recurso valioso para la medicina [58]. Junto con las terapias tradicionales, es capaz de mejorar el tratamiento en casos tan diversos como la disminución del dolor [59], diagnóstico de problemas psicológicos [60], rehabilitación motora [61], tratamiento de exposición para fobias [62] y evaluación de preoperatorio [63]. También sirve a la hora de educar en salud [64]. Esta nueva tecnología no es sólo útil para los pacientes, si no también para los médicos, ayudándoles por ejemplo a tratar de forma más eficiente imágenes médicas [65] así como a integrar la información obtenida a partir de imágenes médicas con el paciente [66], a realizar operaciones tanto con la asistencia de robots [67] como a distancia [68]. Una de sus grandes aplicaciones es el entrenamiento para estudiantes de medicina [69] en distintos tipos de situaciones y procedimientos, normalmente utilizando simuladores de operaciones [70]. Se ha llegado a un punto en el que la realidad virtual no forma parte únicamente de laboratorios de investigación, si no que ya existe en el mercado [71] e incluso algunos hospitales [72]. Se puede por tanto afirmar que la VR se entrelaza cada vez más con el campo de la medicina. Por ello, cualquier avance significativo en esta tecnología posee el potencial de ayudar a mejorar la calidad de vida de los pacientes que la usan actualmente y de permitir que más gente se beneficie de su uso.

3.1. Experimentos previos de supresión perceptual

Existen varios experimentos previos que utilizan la supresión perceptual para dirigir, controlar o predecir la atención del usuario en VR.

El trabajo de Arabadzhiyska et al. [4] (Figura 3.1) utiliza el conocimiento previo sobre las características de las sacadas para crear una nueva técnica de renderizado foveal. En el renderizado foveal se pretende reducir el coste computacional de la escena renderizando a máxima calidad sólo la zona a la que el usuario está mirando directamente. Las zonas a las que no presta atención se renderizan con una calidad menor. El problema surge durante los cambios de atención del usuario. Un cambio de atención o un cambio del lugar al que el usuario está mirando implica que la zona que debe renderizarse con mayor calidad debe moverse con el usuario. Si

este movimiento se produce después del cambio de atención visual, el usuario verá la zona de mala calidad durante unos instantes, lo que produce una mala experiencia. Pero, ¿cómo predecir el nuevo lugar al que acabará mirando el usuario? Con la ayuda de un *eyetracker* y conociendo la naturaleza balística de las sacadas [73], Arabadzhiyska et al. pueden conocer el punto de *aterrizaje* de la mirada del usuario antes de que la sacada termine, por lo que pueden mover la zona de máxima calidad a este punto de aterrizaje antes de que el usuario llege al mismo. Durante las sacadas no somos conscientes de los cambios en nuestro entorno [34] (supresión visual causada por una sacada), por lo que el usuario no se da cuenta del cambio ni ve una zona de mala calidad en ningún momento.

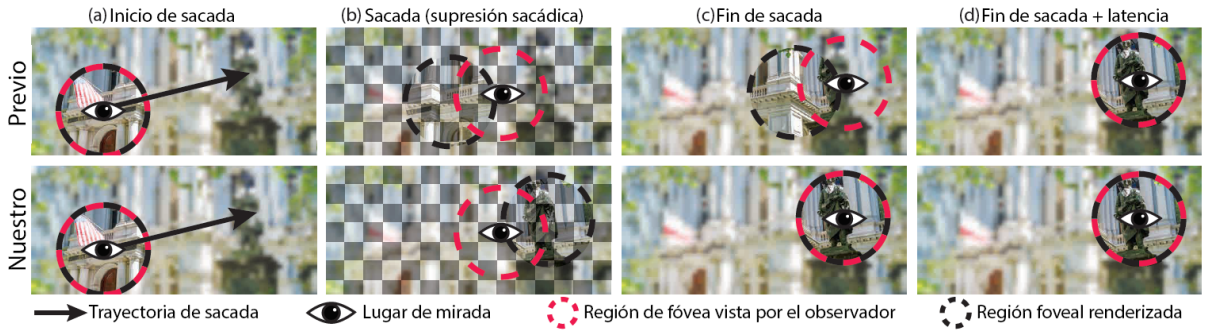
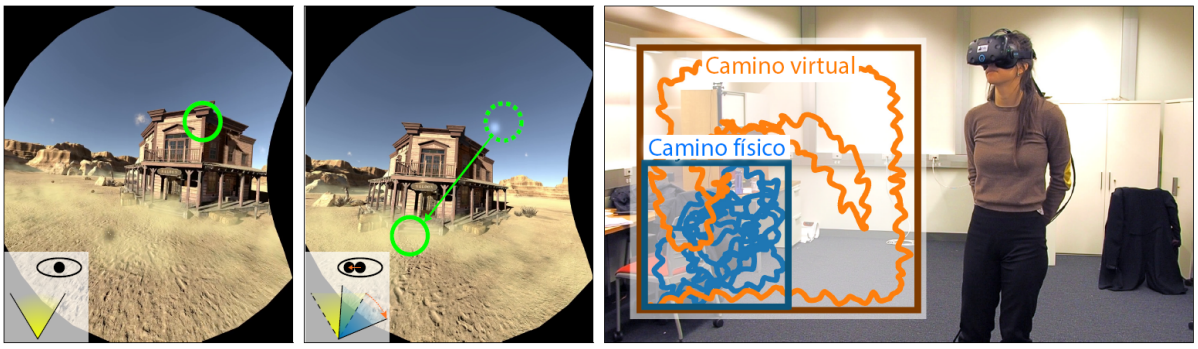


Figura 3.1: Ilustración del método de Arabadzhiyska et al. de renderizado foveal (última columna). Imagen de [4].

En el caso de Sun et al. [3] (Figura 3.2) también se utiliza la supresión visual causada por sacadas. El ámbito de aplicación de esta técnica es la redirección en el espacio físico (o *redirected walking* - RDW). Las técnicas de RDW surgen en VR para intentar solucionar el problema de la diferencia entre el espacio físico disponible para utilizar el HMD y el espacio virtual en el que se puede mover el usuario. Su objetivo es modificar la trayectoria física del usuario sin que este se dé cuenta para que le dé la sensación de que dispone de un espacio mayor, equivalente al virtual, por el que puede moverse libremente. Este tipo de técnicas poseen limitaciones en cuanto al ratio entre espacio físico y virtual que son capaces de relacionar sin que el usuario sea consciente de ello. Llega un punto en el que el usuario ve el mundo virtual demasiado distorsionado o se da cuenta de que sus movimientos reales no se trasladan de la forma que se esperaría al mundo virtual. Para mitigar este problema y permitir un ratio mayor de redireccionamiento, Sun et al. proponen modificar el entorno del usuario cuando este no sea consciente del cambio. Utilizan un *eyetracker* para identificar los momentos en los que se producen las sacadas, y los aprovechan para aumentar la modificación necesaria para la técnica de RDW sin que el usuario note ningún cambio. Además, para aumentar la frecuencia con la que pueden aplicar su técnica, inducen sacadas artificiales mediante estímulos visuales. En este caso la supresión visual se induce con un estímulo visual.

Existen otros trabajos, como el de Langbehn et al. [36] que se aprovechan de la supresión visual que se produce durante los parpadeos para introducir modificaciones en el mundo virtual sin que el usuario se dé cuenta. Bolte et al. [37] también tratan de reorientar y reposicionar al usuario dentro del VE aprovechándose de la supresión sacádica. Estos artículos no serán discutidos en más detalle en este trabajo.

La clave para que estos trabajos funcionen es la supresión perceptual (visual, en este caso) que puede producirse de varias formas, entre ellas durante las sacadas. Su mayor limitación es



(a) Ojo izquierdo, antes de la sacada (b) Después de la sacada (c) Setup experimental. A la izquierda, el camino físico superimpuesto con el camino virtual seguido

Figura 3.2: Ilustración del método de Sun et al. de renderizado sacádico. En la imagen de la derecha se puede observar la diferencia entre el camino recorrido físicamente por el usuario (azul) y el recorrido en la estancia virtual (naranja). Imagen de [3].

la necesidad de un *eyetracker* para poder aplicar estas técnicas. Los *eyetrackers* son dispositivos en general de un precio elevado que no están al alcance del consumidor medio, aunque sí de un equipo de investigación. Uno de los objetivos de este trabajo es conseguir un efecto similar al de estas técnicas eliminando la necesidad de ningún *hardware* adicional. La eliminación del *eyetracker* supone una reducción drástica en el coste del *hardware* necesario (en el caso concreto del *eyetracker* de Pupil Labs para el HMD de HTC Vive, su precio actual es de 1400€ frente a los 1399€ que cuesta el sistema completo de HTC Vive), además de una reducción de carga computacional para la máquina (el procesamiento de la información que llega de las cámaras del *eyetracker* consume al menos la totalidad de un núcleo del ordenador, y requiere de un sistema de almacenamiento con espacio suficiente y lo bastante rápido para la transferencia de las imágenes que se graban). Por ello cabe destacar que la técnica propuesta en este trabajo, que se explicará en la Sección 4.1, no necesita de un *eyetracker* para funcionar. Sin embargo se ha elegido utilizar uno para poder obtener más datos de los participantes.

Para poder eliminar la necesidad de un *eyetracker* es necesario estudiar una forma alternativa y no intrusiva de producir la supresión visual. Se desea evitar utilizar estímulos visuales potencialmente visibles para desencadenar la supresión y depender de una acción consciente del usuario [36]. El método a emplear debería ser no intrusivo para el usuario, y a la vez lo suficientemente fiable para poder aplicarlo de forma consistente.

Se sabe (Sección 2) que el procesamiento de los estímulos auditivos y visuales está relacionado. También que los estímulos auditivos pueden generar sacadas, reduciendo su latencia y mejorando su precisión [74]. En el experimento explicado en la Sección 4.1 se demuestra que se pueden utilizar distintos tipos de sonidos para desencadenar supresión visual de forma consistente, fenómeno que podría ser aplicado a las técnicas descritas en esta sección que se sirven de la supresión perceptual para su funcionamiento. Este efecto no depende de un *eyetracker* para su funcionamiento, por lo que cumple con los requisitos propuestos en este trabajo.

4. Diseño del experimento

En este capítulo se describe la idea que fundamenta este trabajo: la búsqueda de una nueva técnica de supresión perceptual (en concreto visual) para VR desencadenada mediante estímulos sonoros, a partir de la cual se pueda construir un modo de controlar la atención, con posibles aplicaciones en el campo de la medicina. Los resultados obtenidos se analizan en el Capítulo 5.

4.1. Descripción del experimento

Las interacciones entre el sistema humano visual y el auditivo son complejas y todavía no se entiende del todo su funcionamiento a nivel del sistema nervioso. Sin embargo, está demostrado que un estímulo de una modalidad puede influir en la otra, provocando tanto una mejora de su rendimiento, bien conocida [74], como un efecto de supresión [75, 76, 24].

El objetivo de este experimento es encontrar un sonido que desencadene sacadas o produzca algún otro tipo de supresión visual. Como se ha visto en la Sección 3.1, una vez se obtiene un método de supresión perceptual éste se puede aplicar a varias técnicas de control de la atención. La supresión en sí hace que el usuario no sea consciente de las modificaciones del VE a su alrededor, por lo que es una forma por sí misma de controlar la atención, o desviarla, en el momento más oportuno.

Experimento realizado

Participantes. Un total de 35 participantes realizaron el experimento. La media de edad de los participantes era de 24 años con una desviación de 8. Del total, 13 eran mujeres. Todos tenían visión normal o corregida y no presentaban problemas auditivos. 34 participantes no conocían el objetivo del experimento.

Equipamiento. El *hardware* utilizado durante el experimento ha sido un sistema de RV completo HTC Vive Pro (las gafas de VR (HMD) y dos *trackers* con una superficie calibrada de 4x1.5m por la que el usuario podía moverse libremente durante el experimento). Integrado en el sistema se encuentra un *eyetracker* de Pupil Labs (frecuencia de muestreo = 120Hz). Las especificaciones completas del HMD y del *eyetracker* se pueden consultar en la Sección 2.2. Un único ordenador se utilizaba para el experimento, con un procesador i7-7700 a 3.6GHz, 16GB de RAM y una tarjeta gráfica Nvidia 1060GTX con 6GB de memoria DDR5 dedicados. En cuanto al *software*, todas las escenas fueron programadas en Unity 3D (versión 2018), utilizando los

plug-in de Pupil Labs para grabar, el *software* de *Steam VR* para integrar el sistema de VR y el *plug-in* de captura *VR-capture* disponible para Unity. El SO utilizado es Windows 10.

Estímulos visuales. Los estímulos visuales (Figura 4.1) consistían en cinco formas simples (círculo, cuadrado, rombo, pentágono y estrella de cinco puntas) con relleno blanco y un borde de un grosor del 5 % del tamaño de la forma de color gris para evitar que el estímulo se pudiera confundir con un fondo blanco. El tamaño de los estímulos es de un grado en el campo visual. Al aparecer, los estímulos visuales permanecen 24ms en el campo de visión del usuario. Tanto el tamaño del estímulo como el tiempo de visualización se han fijado siguiendo el trabajo de Hidaka et al. [76] en el que se demuestra que los estímulos auditivos pueden empeorar la discriminación de estímulos visuales en una pantalla convencional. El estímulo visual puede aparecer en tres localizaciones posibles de forma aleatoria: en el centro del campo visual del usuario (FOV), cuatro grados a la derecha del centro del FOV o cuatro grados a la izquierda del centro del FOV. En adelante, estas tres condiciones serán *visFront*, *visRight* y *visLeft* respectivamente. Antes de la realización de este experimento se realizó un experimento previo con 7 participantes para asegurar que estos estímulos son visibles en unas condiciones idénticas excepto por la ausencia de estímulos sonoros. En el Anexo A puede encontrarse una descripción detallada de este experimento y sus resultados y la motivación para las distintas localizaciones de los estímulos visuales.

Figura 4.1: Estímulos visuales presentados durante el experimento.

Estímulos sonoros. Los estímulos auditivos consisten en un conjunto de seis posibles sonidos, elegidos en base a la literatura: ruido blanco [76], ruido rosa [77], ruido marrón (los cambios aleatorios entre una nota y la siguiente pueden llamar la atención más que algo predecible [21]), frecuencia pura [21], sonido de supervivencia [21] y voz humana [21]. La hipótesis es que estos sonidos llamarán la atención del participante, que producirá sacadas o algún otro tipo de respuesta visual hacia el origen del sonido. Todos los estímulos presentan la misma duración, 400ms, para que los sonidos más complejos (el de supervivencia y la voz humana) sean distinguibles. Los sonidos aparecen en tres posibles localizaciones, siempre fuera del FOV del participante: detrás, a la izquierda y a la derecha. Los sonidos se localizan fuera del FOV del participante para desencadenar una mayor desviación de su atención [21]. En la Tabla 4.1 se puede observar una relación completa de los estímulos presentados en este experimento.

Procedimiento. Los participantes se encuentran en una sala virtual similar a un salón a tamaño real como el de la Figura 4.2. Aparecen detrás de la mesita y la alfombra y tienen un espacio real de 4x1.5m para moverse libremente durante el experimento. Antes de introducirles en esa escena se les muestra otra similar pero sin mobiliario, en la que se explican las mecánicas del experimento que realizarán y pueden acostumbrarse al VE. Al participante se le explica que verá aparecer formas visuales sencillas en su campo visual y que cuando vea una deberá avisar al experimentador. El participante no sabe cuántas formas diferentes pueden aparecer, ni cuales, antes de verlas. También se le dice que escuchará sonidos aleatorios durante el experimento que pueden o no coincidir con la aparición de los estímulos visuales, pero que sólo debe avisar al experimentador cuando detecte un estímulo visual, aunque reconozca su forma. Cuando el experimentador es avisado pulsa un botón del teclado y aparece una pantalla traslúcida en las

Factores *vs* Condiciones	**Condición visual** 18 estímulos	**Condición sonora** 18 estímulos	**Condición bimodal** 18 estímulos
Factor posición (visual)	3 niveles (aleatorio)	0	3 niveles (aleatorio)
Factor forma (visual)	5 niveles (uniforme)	0	5 niveles (aleatorio)
Factor tipo sonido	0	6 niveles (aleatorio)	6 niveles (uniforme)
Factor localización sonido	0	3 niveles (aleatorio)	3 niveles (uniforme)

Tabla 4.1: Tabla resumen de las condiciones exploradas en este experimento.

cuatro paredes de la habitación con la pregunta *¿Qué ha visto?*. En ese momento el experimentador pregunta al participante cuál es el estímulo que ha visto para registrar su respuesta. El usuario contesta con el nombre de la forma que ha visto si la ha reconocido, o dice que no la ha reconocido si ese es el caso. El experimentador guarda la respuesta del usuario pulsando otro botón del teclado (uno distinto para cada estímulo visual, otro para cuando el usuario ha visto algo pero no ha reconocido la forma) y el experimento continúa. Si el experimentador ha pulsado el primer botón del teclado por error sin que el participante le haya indicado nada, existe otra tecla para indicar esta posibilidad y anular el registro de la respuesta.

Durante el experimento pueden aparecer estímulos sólo visuales, audiovisuales (la forma precedida por un sonido en una localización fuera del campo visual del participante), o sólo auditivos, en los que el usuario no debe avisar al experimentador. Llamaremos a estas condiciones *condVis*, *condBi* y *condSoun* respectivamente. En cada experimento, el participante ve 18 estímulos de cada una de las tres condiciones. Durante el diseño del experimento, se decidió utilizar este procedimiento por varias razones.

- Los estímulos de la condición *condVis* sirven como centinelas para asegurar que el usuario está prestando atención a la tarea indicada.
- Los estímulos *condVis* permiten obtener una segunda confirmación de su visibilidad para el participante. Además sirven como línea de base para compararlos con los estímulos audiovisuales, lo que permite tanto un estudio estadístico de los datos entre distintos usuarios como dentro de un mismo usuario.
- Al haber estímulos de distintos tipos, es menos probable que el participante asocie el sonido con la aparición del estímulo visual. Si esto ocurriera, el participante podría anticipar el momento de aparición de un estímulo visual al escuchar el sonido, y se desea evitar este efecto.
- Los estímulos *condSoun*, poseen una función distractora para evitar el efecto de habituación ya mencionado.
- Los estímulos *condBi* son la forma más directa de saber si el participante es consciente de que se ha producido un cambio visual cuando escucha el sonido. Si el usuario no reporta haber visto una forma, esto quiere decir que la parte sonora del estímulo ha distraído la

atención del participante el tiempo suficiente como para que este no sea consciente de haber visto el estímulo visual durante la duración del mismo.

Cada estímulo visual aparece durante 24ms, cada sonido dura 400ms, y el tiempo entre estímulos tiene una duración aleatoria entre 5 y 10s. En el caso de los estímulos *condBi*, la parte visual aparece 100ms después de que el sonido haya comenzado a reproducirse. El orden de los estímulos es aleatorio para cada participante para evitar efectos debidos al orden de presentación. Durante el experimento, los usuarios oyen los 6 sonidos una vez en cada una de las 3 localizaciones posibles (18 estímulos *condBi*). También ven 18 estímulos *condVis* repartidos de tal forma que cada geometría se repite al menos 3 veces. Ya que en el experimento previo (Anexo A) se ha comprobado la visibilidad de los estímulos visuales en las tres localizaciones posibles dentro del FOV del participante, en este experimento la localización de la parte visual del estímulo tanto en la condición *condBi* como en *condVis* es aleatoria y no se muestrea de forma exhaustiva para mantener el tamaño del experimento y sobre todo su duración dentro de unos límites aceptables que no causen malestar debido a pasar un tiempo excesivo en VR. Los 18 estímulos restantes de la condición *condSoun* se presentan también de forma aleatoria.

Todos los usuarios ven los mismos estímulos del experimento (*condBi*) y control (*condVis*) iguales, y se les presenta el mismo número de estímulos (18x3 condiciones, 54). Adicionalmente los participantes escuchan un sonido de fondo durante la duración del experimento, que consiste en sonido ambiente de un parque [78] que se escucha a través de una de las ventanas del salón, que está abierta, y un podcast de noticias [79] que se escucha por un altavoz situado en la escena a la derecha de la televisión. La intención del sonido de fondo es hacer la escena más compleja y realista, y evitar que el único sonido que escuche el participante sea el de los estímulos *condSoun*.

Antes de comenzar, el usuario recibe una explicación del procedimiento del experimento y el sistema para comunicar las respuestas al experimentador. Se le informa de la presencia de sonidos aleatorios durante el experimento, de que los estímulos visuales aparecerán y desaparecerán muy rápidamente y de que debe estar centrado en reconocer la aparición de los estímulos visuales. El usuario es informado de que si siente cualquier tipo de malestar o mareo debe avisar al experimentador para que pare el experimento. Antes de que el experimentador le ponga el HMD, el participante rellena un cuestionario con datos sociodemográficos (edad, género, estudios, trabajo, problemas visuales y uso y experiencia previa con VR). Una vez puesto el HMD, se realiza una calibración del *eyetracker* en la que el usuario debe mirar fijamente a un punto que se mueve por siete localizaciones fijas de la pantalla. Por último, después de realizar el experimento, el usuario rellena la segunda parte del cuestionario (si ha sentido algún tipo de malestar, si ha visto u oído algo extraño, o ha notado algún cambio a lo largo del tiempo en el experimento, si ha sido consciente de los temas que se trataban en el podcast que se escuchaba de fondo y un apartado de escritura libre por si quiere comunicar algo más al experimentador). En el Anexo B se puede ver el cuestionario que rellenaron todos los participantes. Opcionalmente, como agradecimiento por su participación, los usuarios podían probar un minijuego del HMD después de realizar el experimento.

Como última consideración sobre el experimento, es necesario tener en mente que los estímulos elegidos en esta ocasión están desprovistos de cualquier tipo de semántica en relación con la escena presentada. Esto implica que los estímulos visuales (formas geométricas simples, planas, que flotan en el aire) destacan de forma intencionada en el contexto en el que han sido colocados (un escenario realista y complejo). Lo mismo ocurre con los sonidos, que no están

Figura 4.2: Escena obtenida de la tienda de *Assets* de Unity 3D [80], en la que los usuarios realizan el experimento. Vista del HMD. La posición inicial de los participantes se encuentra a la derecha de la mesa, en frente de la televisión.

integrados en el ruido de ambiente. Este hecho no es casual, si no calculado, en un intento de que el efecto encontrado en este experimento sea lo más conservador posible. Si los estímulos visuales hubieran tenido un significado semántico relacionado con la escena vista por los participantes del experimento, estos habrían pasado desapercibidos más fácilmente, potenciando el efecto de supresión perceptual. De la misma manera, si los estímulos sonoros formasen parte de la ambientación, es posible que los participantes no hubiesen sido conscientes de la presencia de los mismos. Ambos estímulos pueden adaptarse al contenido en todo momento, para acentuar el efecto aquí mostrado.

5. Análisis de los resultados obtenidos

En esta Sección se realiza un análisis de los resultados obtenidos en el experimento de la Sección 4.1. Primero se comprueba que existen diferencias significativas entre las condiciones *condVis* y *condBi* (Sección 5.1) y se comparan los resultados de la condición visual con el experimento previo (5.2). Una vez comprobado que existen estas diferencias, se estudia cuáles de los factores manipulados en el experimento influyen de forma significativa en ellas (Sección 5.3). Se analiza la detección y el reconocimiento de estímulos visuales por separado, y se profundiza más en las particularidades de los estímulos de la condición *condBi* (Sección 5.3). Por último, se comentan los resultados obtenidos con los *eyetrackers* de forma cualitativa (Sección 5.5) y parte de la información subjetiva reportada por los participantes sobre la intrusividad de los sonidos presentados en el experimento (Sección 5.4). En la Sección 5.6 se resumen los resultados obtenidos en el análisis.

5.1. Análisis de las condiciones visual y bimodal

En total, se presentan 36 estímulos visuales a cada participante. Estos 36 estímulos visuales son siempre una de las cinco formas descritas en el Capítulo 4. De esas 36 formas, la mitad (18) aparecen 100ms después de que se haya iniciado uno de los seis sonidos utilizados en el experimento.

Siete participantes fueron eliminados del análisis por no superar el 25 % ni de detección ni de reconocimiento de los estímulos visuales. Para esta eliminación sólo se han utilizado los estímulos de la condición *condVis* que servían como control en el experimento. Los resultados presentados en esta Sección corresponden a los 28 participantes restantes.

Se presenta un resumen de los resultados en la Tabla 5.1 para la detección de los estímulos visuales y en la Tabla 5.2 para el reconocimiento. En las esta Sección y en la Sección 5.3 se explica cómo se han realizado los análisis que aparecen en ellas.

Si se tiene en cuenta ese total de 36 formas (*condVis*+*condBi*), la detección media de los participantes es del 36.44 % (±6.97 % desviación típica), de los cuales reconocen correctamente un 58.04 % (±14.55 %). Debido a la diferencia con la media del experimento previo (Anexo A), en el que se detectaban y reconocían 50 estímulos visuales en igualdad de condiciones con el experimento de la Sección 4.1) en el que el porcentaje de detección era aproximadamente del 88 % y el de reconocimiento del 72 % se sospecha que existen diferencias significativas entre las condiciones *condVis* y *condBi*.

Detección de estímulos visuales	**Influye**	**P-valor**	**Figura asociada**	**Efecto**
Condición (*condVis* vs *condBi*)	**Sí**	$< 0{,}001$	Figura 5.1	Decrece para *condBi*. La presencia de sonido disminuye la detección de los estímulos visuales
Forma del estímulo Visual	No	0.821	Ninguna	La forma no influye en la detección
Localización del sonido	No	0.199	Figura 5.5	Las tres localizaciones producen una disminución similar en la detección
Tipo de sonido	No	0.57	Figura 5.6	Los seis sonidos explorados producen una disminución similar en la detección para *condVis+condBi*.

Tabla 5.1: Tabla resumen de los resultados obtenidos para la detección de estímulos visuales en el experimento de la Sección 4.1

Reconocimiento de estímulos visuales	**Influye**	**P-valor**	**Figura asociada**	**Efecto**
Condición (*condVis* vs *condBi*)	**Sí**	$< 0{,}001$	Figura 5.3	Decrece para *condBi*. La presencia de sonido disminuye el reconocimiento de los estímulos visuales
Forma del estímulo Visual	**Sí**	$< 0{,}001$	Figura 5.3	El pentágono es significativamente diferente del resto. Para *condBi* la influencia significativa de este factor desaparece.
Localización del sonido	**Sí**	$< 0{,}001$	Figura 5.5	La presencia del sonido en una de las tres localizaciones estudiadas disminuye significativamente la tasa de reconocimiento
Tipo de sonido	**Sí**	0.015	Figura 5.6	La presencia de uno de los seis tipos de sonido estudiadios disminuye significativamente la tasa de reconocimiento

Tabla 5.2: Tabla resumen de los resultados obtenidos para el reconocimiento de estímulos visuales en el experimento de la Sección 4.1

Para comprobarlo se realiza un test de Wilcoxon (puesto que no podemos asegurar la normalidad de nuestros datos) para contrastar si las medias de detección de los estímulos visuales en la condición *condVis* y en la condición *condBi* pertenecen a la misma distribución. Se rechaza la hipótesis nula con un p-valor $< 0{,}001$. Para el caso de reconocimiento en ambas condiciones, también se rechaza la hipótesis nula con un p-valor $< 0{,}001$. Cabe destacar que el test de Wilcoxon, al no asumir la normalidad de las distribuciones presentadas es más estricto a la hora de rechazar la hipótesis nula, por lo que el enfoque utilizado es conservador.

Se puede por tanto concluir que existe una diferencia significativa tanto en la detección como en el reconocimiento de los estímulos presentados ante los participantes dependiendo de la condición a la que pertenezca el estímulo (*condBi* o *condVis*). En concreto **la tasa de reconocimiento y detección decrece para la condición bimodal** (detección del 18.25 % (±18.75 %) y reconocimiento del 5 % (±11.14 %) del total de estímulos) con respecto a la condición visual (detección del 69.21 % (±36.41 %) y reconocimiento del 55.76 % (±32.44 %) del total de estímulos), lo cual nos indica que existe una interacción significativa en la que el sonido influye en la percepción visual. Existen varias hipótesis que explican este comportamiento:

- Se ha demostrado que los sonidos pueden producir sacadas [81]. Es posible que los participantes realicen sacadas hacia el origen del sonido (fuera de su FOV, por lo que la sacada debería dirigirse hacia la periferia de su campo de visión) sin ser conscientes de ello. La sacada debe ocurrir en algún momento durante los primeros 100ms del sonido y mantenerse durante la aparición y desaparición del estímulo visual (que dura 24ms), lo que hace que el participante no pueda ver la forma que se presenta ante él en la zona foveal de su campo de visión.
- Como se explica en la Sección 2.1 existen zonas del sistema nervioso humano que se ven afectadas por las modalidades visual y auditiva. En el Capítulo 3 se explica que existen efectos de supresión visual causados por estímulos sonoros. Es posible que el procesamiento del estímulo sonoro, que aparece en congruencia temporal con el visual y se extiende más allá de su duración, suprima el procesamiento de la información visual en el cerebro.

5.2. Comparación con el experimento previo

Se realiza un test de Wilcoxon para comparar la detección y el reconocimiento de la condición visual de este experimento (*condVis*) con los resultados del experimento previo. En el experimento previo (explicado en el Anexo A), se detectaban y reconocían 50 estímulos visuales. En el experimento previo se demuestra que los dos factores explorados (forma y localización del estímulo visual) no influyen en la detección de los estímulos visuales, y que únicamente la forma influye en el reconocimiento de los mismos. El experimento previo demuestra la visibilidad de estos estímulos, con una tasa de detección media del 88.10 % y del 71.96 % de reconocimiento.

El reconocimiento obtenido en este experimento no es significativamente diferente del obtenido en el experimento previo. Sin embargo, el porcentaje de detección sí que es significativamente menor. La diferencia puede ser debida a varias razones:

- Los usuarios que han realizado este experimento no son los mismos que realizaron el experimento previo.

- La presencia de distractores (los estímulos sonoros) puede afectar al comportamiento de exploración y búsqueda del usuario negativamente.
- Mientras que en el experimento previo todos los participantes tenían algún tipo de experiencia previa con la VR, en este experimento solo 18 de los participantes habían probado la VR alguna vez, por lo que es posible que la novedad del experimento les distrajera de su tarea de búsqueda.

Debido a esta diferencia, el análisis estadístico se centrará en comparar las condiciones *condVis* y *condBi* de este experimento, y no se comparará la condición audiovisual con el experimento previo.

5.3. Análisis de significancia de los factores

Se ha utilizado un *Generalized linear mixed model* (GLMM) para realizar un análisis estadístico sobre la influencia que los distintos factores del experimento tienen sobre la respuesta de los participantes. Se ha utilizado este modelo y no un análisis ANOVA porque la variable objetivo no es normal. Además, queremos tener en cuenta efectos aleatorios (en este caso el efecto del usuario). Por último, el modelo debe ser robusto y poder manejar situaciones en las que no todos los usuarios han visto los mismos estímulos debido tanto al efecto de la supresión visual como a otros factores discutidos en el Anexo A. Después del análisis de significancia, se realiza un análisis *post-hoc* con tests *pairwise* de Fisher, ya que este test permite comparar variables binarias, para encontrar cuáles de los niveles son significativos dentro de cada factor con significancia.

No se estudia la influencia de la posición del estímulo visual, ya que en el experimento previo (Anexo A) se demuestra que los cambios en este factor no son significativos a la hora de detectar o reconocer el estímulo visual. Los factores que se estudian son, por tanto: la forma del estímulo visual (5 niveles), la localización del sonido (3 niveles), el tipo de sonido (6 niveles) y el usuario como efecto aleatorio. La variable respuesta se ha binarizado y es modelada como una distribución binomial.

En ninguno de los análisis realizados el efecto aleatorio debido al usuario que ha realizado el experimento ha resultado ser influyente en la respuesta obtenida.

El análisis se ha separado en dos niveles. Primero, se ha analizado la influencia de los factores en la detección del estímulo visual. Después, se ha analizado la influencia de los mismos en el reconocimiento del estímulo visual.

Detección. En un primer paso, se realiza un análisis sobre las respuestas a todos los estímulos con componente visual (*condVis*+*condBi*). La presencia del sonido tiene un efecto significativo (p-valor $< 0{,}001$) sobre la variable respuesta (Figura 5.2). Cualitativamente este efecto se observa en la Figura 5.1, en la que se puede observar la media de detección para cada usuario separada por condición. El porcentaje de detección en la condición *condBi* es consistentemente inferior al de la condición *condVis* en todos los usuarios, indicando que **el sonido tiene una fuerte influencia en el hecho de si un usuario ve o no un estímulo**. Incluso hay tres usuarios (uno de ellos con un porcentaje de detección de estímulos en la condición *condVis* del

100 %) que no han sido capaces de detectar ninguno de los estímulos en la condición *condBi*.

En un segundo paso, se analizan únicamente las respuestas correspondientes a la condición *condBi* para estudiar si dentro de la misma existen diferencias en cuanto a los distintos niveles de los factores que afectan al estímulo auditivo, sin encontrar ningún efecto significativo). No se encuentra ninguna diferencia significativa dentro de esta condición, por lo que se puede concluir que tanto las localizaciones utilizadas como fuente del sonido como los distintos tipos de sonidos afectan de una forma similar a la detección de estímulos visuales, por lo que **debe existir un mecanismo común a estos sonidos que sea el causante de la supresión visual**.

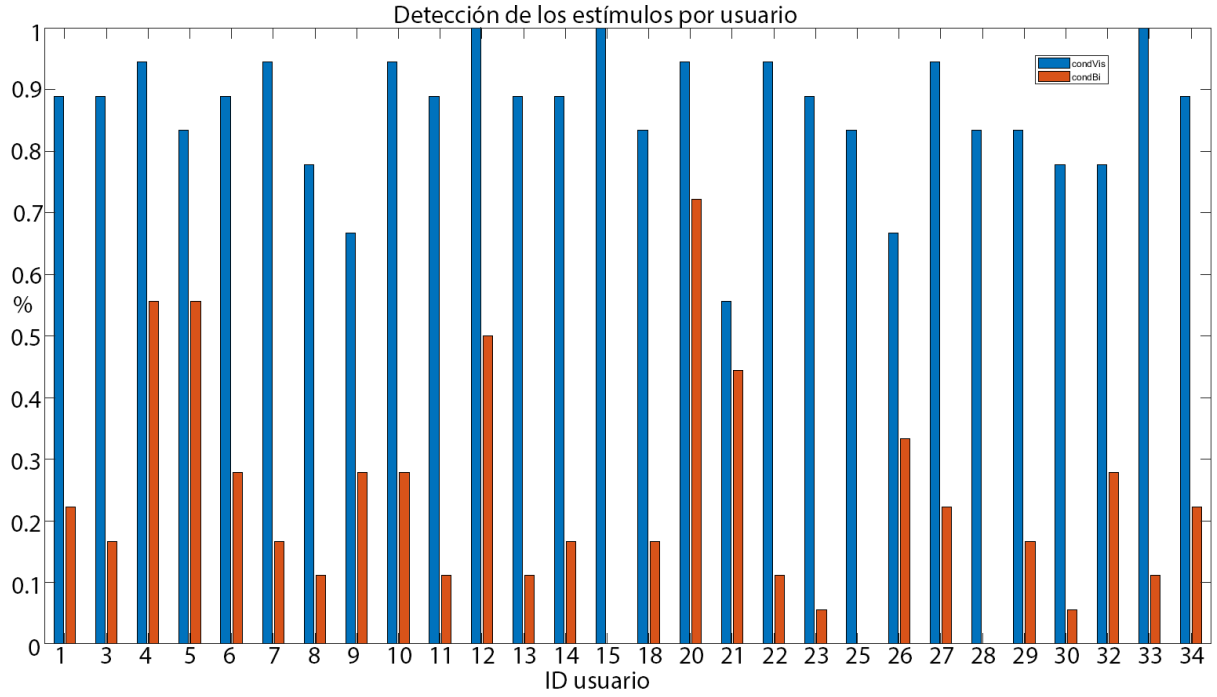


Figura 5.1: Media de detección de cada uno de los usuarios que han realizado el experimento. Azul: *condVis*. Naranja: *condBi*.

Reconocimiento. Se realiza un análisis sobre la proporción de estímulos visuales reconocidos correctamente de entre aquellos que han sido detectados. La presencia del sonido influye en el hecho de que el usuario sea capaz de reconocer correctamente el estímulo (Figura 5.4). La forma del estímulo visual (como se ha demostrado en el experimento del Anexo A) sigue influyendo en el reconocimiento del mismo, sin embargo su influencia pasa a ser no significativa si se tienen en cuenta únicamente los estímulos de la condición *condBi*. El efecto de este factor separado por condición puede observarse en la Figura 5.3. El reconocimiento de estímulos es consistentemente mayor en la condición *condVis*. Sin embargo, existen diferencias que ocurren dentro de ambas condiciones, como la mayor tasa de reconocimiento de la forma *extrella*. Pese a que la capacidad de reconocimiento se ve mermada en los estímulos bimodales, se puede apreciar cierta consistencia en el reconocimiento de ambas condiciones lo que podría indicar un procesamiento visual independiente del efecto de supresión producido. **Es posible que los estímulos visuales se sigan procesando aunque el usuario no sea consciente de ello**.

Influencia del sonido dentro de la condición bimodal. Se han estudiado los dos factores relacionados con el sonido únicamente en los estímulos de la condición *condBi*. Ni la localiza-

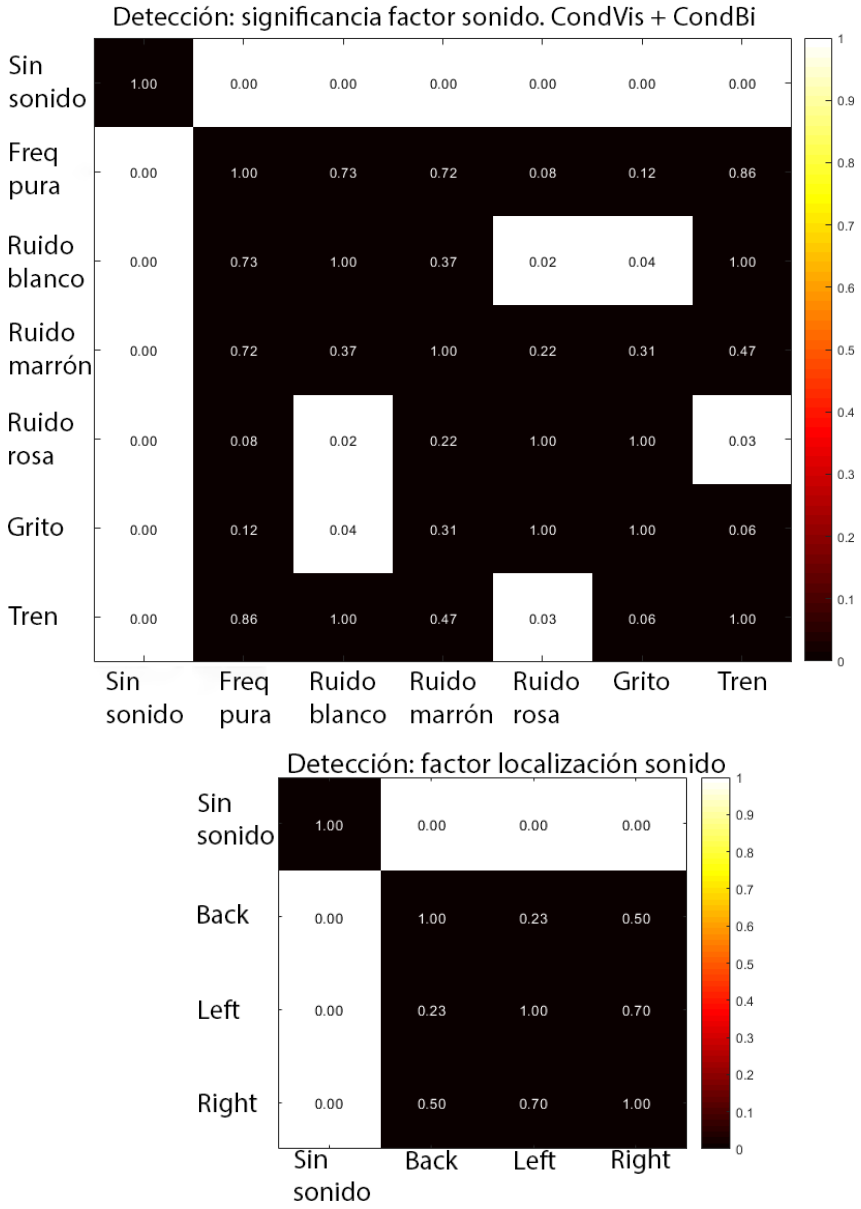


Figura 5.2: Análisis *post-hoc* realizados tras ver que la presencia de sonido influye de forma significativa en la detección de los estímulos visuales. Arriba, tipo de sonido. Abajo, localización del sonido. En blanco los p-valores significativos ($p < 0{,}05$). Como puede observarse, el efecto de la ausencia de sonido es significativamente distinto al resto. El efecto es una disminución de la detección.

ción (p-valor 0.199) ni el tipo de sonido (p-valor 0.573) ni su interacción (p-valor 0.308) tienen influencia significativa en la detección de estímulos. Es decir, **es la presencia o ausencia del sonido lo que causa la supresión visual**. En cambio, tanto la localización del sonido (p-valor $< 0{,}001$) como la interacción entre esta y el tipo de sonido (p-valor 0.004), así como el tipo de sonido (p-valor 0.015), sí que influyen a la hora de reconocer el estímulo.

En la Figura 5.5 se pueden observar los porcentajes de detección y reconocimiento para cada una de las tres localizaciones del sonido. *El porcentaje de reconocimiento es menor cuando el*

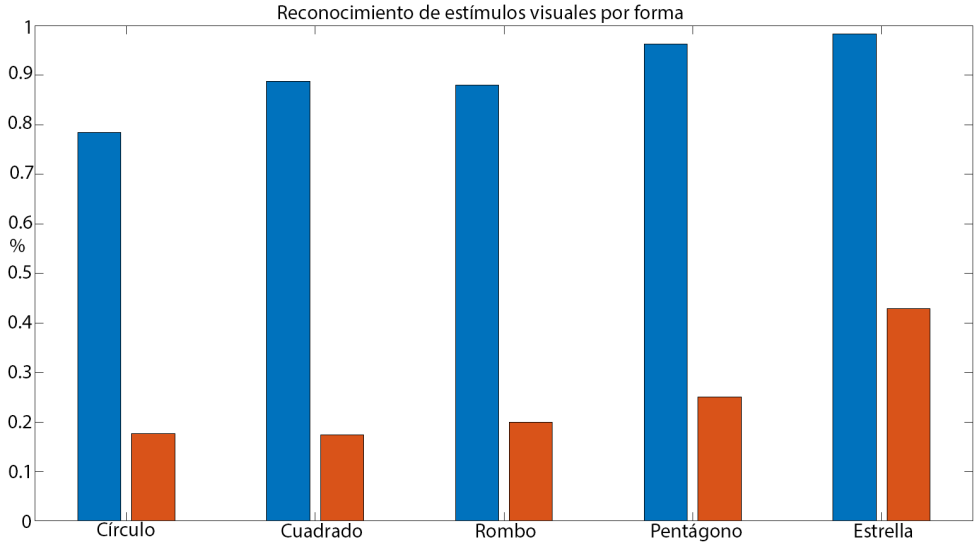


Figura 5.3: Tasa de reconocimiento por forma comparando las dos modalidades. Azul: *condVis*. Naranja: *condBi*.

sonido proviene directamente desde atrás, aunque este efecto no es signifivativo. La Figura 5.6 muestra los porcentajes de detección y reconocimiento según los distintos tipos de sonido de la condición bimodal. En este caso, tampoco se encuentras diferencias significativas en los tests *post-hoc* dentro de la condición bimodal. Aún así, se ven efectos fuertes como el hecho de que *ningún usuario ha reconocido correctamente ningún estímulo visual cuando el sonido asociado es el ruido rosa*. El *ruido blanco* también presenta un porcentaje de reconocimiento por debajo del 10 %, lo que concuerda con el experimento de Hidaka et al. [76]. Por último, la frecuencia pura también presenta un porcentaje de reconocimiento bajo, lo que apoya la idea de Spence et al. [21] de que las frecuencias puras producen un mayor desvío de la atención, al ser sonidos que no solemos escuchar en un entorno natural.

5.4. Información subjetiva proporcionada por los usuarios

Después del experimento, los usuarios rellenaron un formulario en el que se les preguntaba específicamente si habían oído algo extraño entre otras cosas (ver el Anexo B para consultar las preguntas del cuestionario). A continuación se resumen las respuestas proporcionadas en relación con los estímulos auditivos:

- Cuatro de los 35 usuarios indicaron que los sonidos no parecían corresponder con la escena.
- Tres usuarios comunicaron que había ruidos molestos en el experimento. En concreto, los pájaros que piaban por la ventana, las interferencias (ruido blanco, rosa o marrón) que se escuchaban y los gritos, respectivamente.
- Los sonidos que más llamaron la atención fueron los gritos (14 usuarios) y el sonido del tren (7 usuarios).

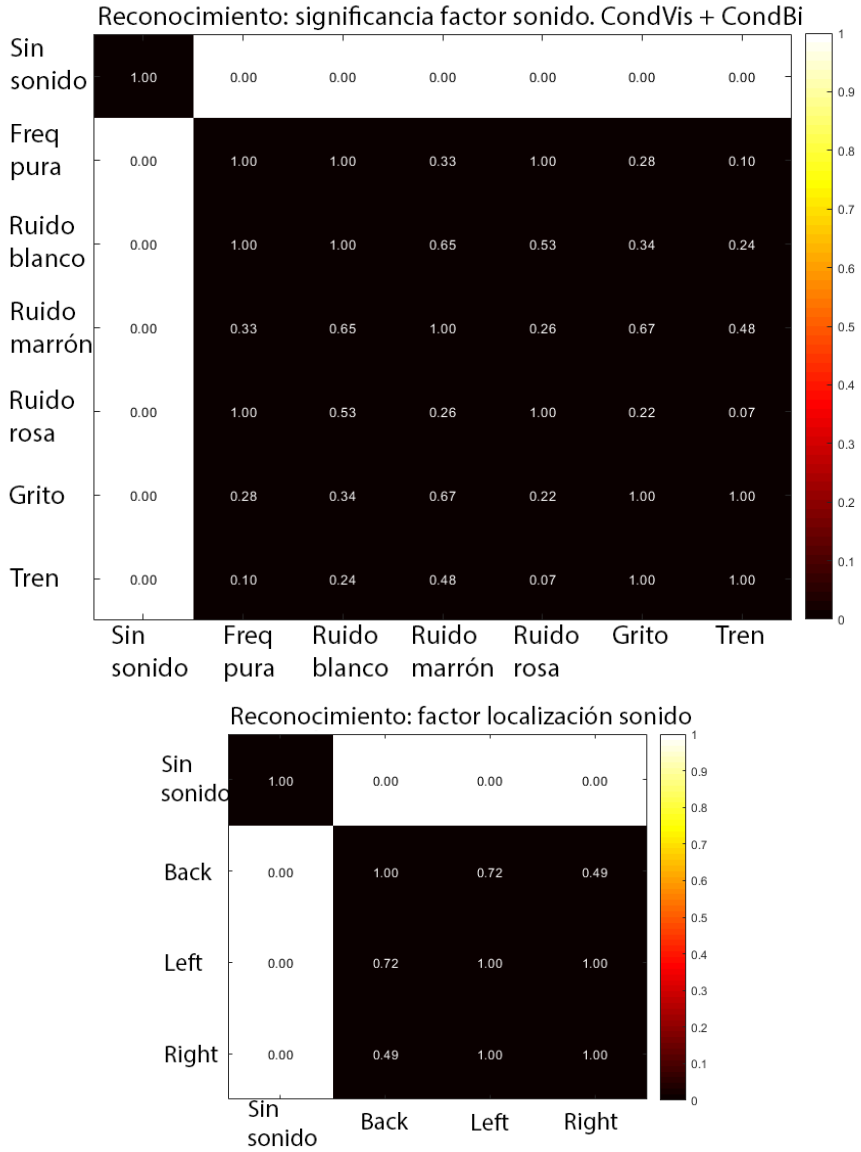


Figura 5.4: Análisis *post-hoc* realizados tras ver que la presencia de sonido influye de forma significativa en el reconocimiento de los estímulos visuales. Arriba, tipo de sonido. Abajo, localización de sonido. En blanco los p-valores significativos ($p < 0{,}05$). Como puede observarse,el efecto de la ausencia de sonido es significativamente distinto al resto. El efecto es una disminución del reconocimiento.

En general parece que los sonidos utilizados no son molestos para la mayoría de los participantes, pero parece que los sonidos más complejos (el grito y el tren) sí que son más intrusivos en la experiencia. Teniendo esta información en cuenta se pueden diseñar sonidos que causen supresión visual y que pasen desapercibidos para los usuarios dentro del sonido ambiente. Pese a estas observaciones, no hay diferencias significativas en los distintos tipos de sonido utilizados en cuando a su influencia sobre la disminución del porcentaje de aciertos de los participantes, por lo que todos han sido eficaces para este experimento.

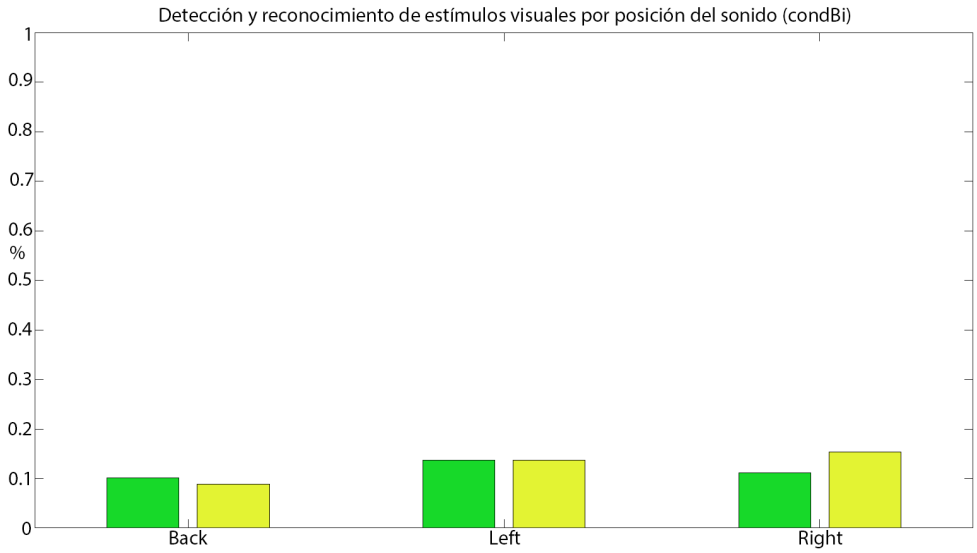


Figura 5.5: Tasa de reconocimiento según la localización del sonido. Verde: Detección en la condición *condBi*. Amarillo: reconocimiento en la condición*condBi*.

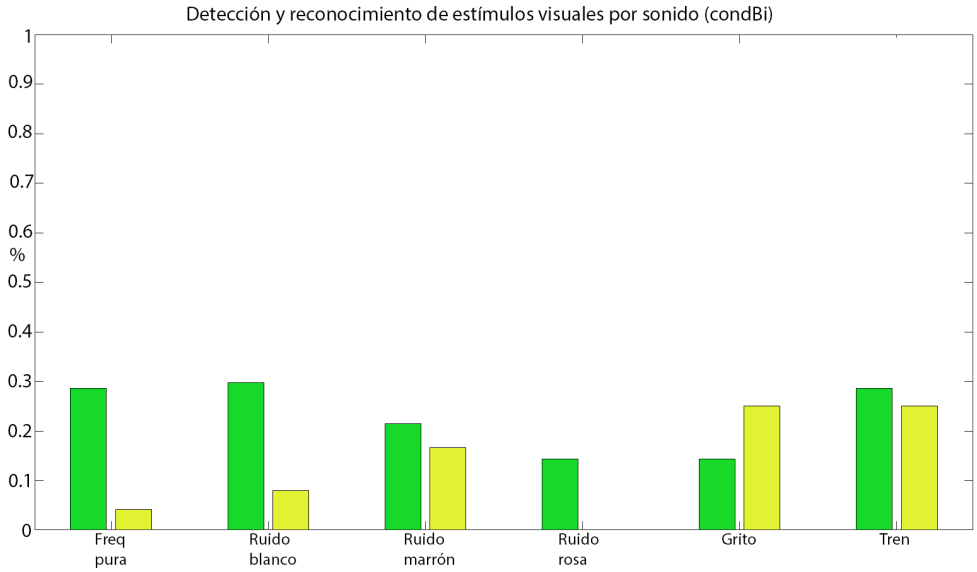


Figura 5.6: Tasa de reconocimiento según el tipo de sonido. Verde: Detección en la condición *condBi*. Amarillo: reconocimiento en la condición *condBi*.

5.5. Análisis cualitativo de los datos de eyetracker

Como se explica en la Sección 3.1, el experimento descrito en este trabajo no necesita el uso de un eyetracker. Sin embargo, se ha utilizado uno para obtener más información de los participantes. En este apartado se presenta un primer análisis cualitativo de los datos de uno de los participantes que permite obtener ciertas intuiciones sobre el comportamiento del sistema

visual durante la supresión perceptual. Este conocimiento servirá como base para realizar en el futuro un análisis completo y cuantitativo de los datos obtenidos. En la Figura 5.7 se puede observar un ejemplo de lo que ocurre cuando al usuario se le presenta un sonido. Como se suponía, una de las causas de que el usuario no vea el estímulo visual es la supresión sacádica inducida por el sonido. Esto concuerda con los trabajos que indican que el sonido puede producir sacadas [74, 82]. Sin embargo, es necesario un análisis más profundo, metódico y formal para llegar a afirmar que esta es la causa única y principal del efecto de supresión visual observado.

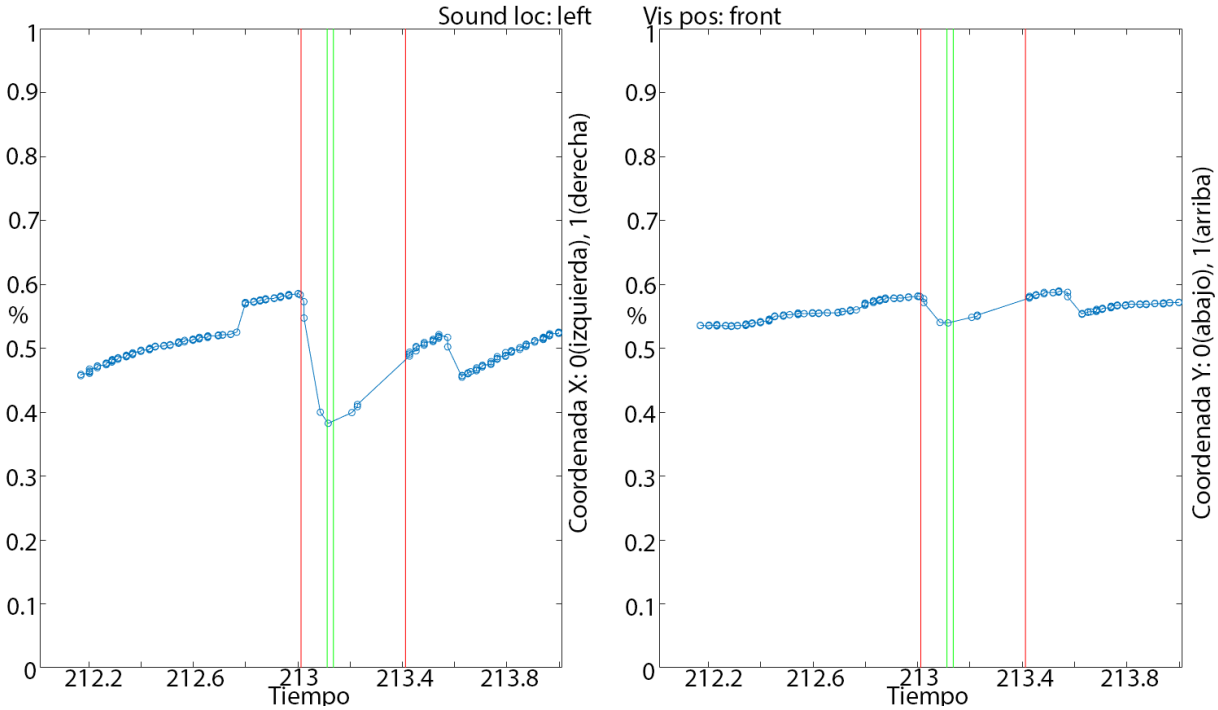


Figura 5.7: Izquierda: Posición horizontal de la mirada a lo largo del tiempo. Derecha: Posición vertical de la mirada a lo largo del tiempo. Las líneas rojas verticales indican la duración del sonido, y las verdes la del estímulo visual. En este ejemplo en concreto, el sonido se localizaba a la izquierda del usuario y el estímulo visual directamente en el centro de su FOV. Se produce una sacada en dirección del sonido, alejando la vista del estímulo visual, que desaparece antes de que el ojo vuelva a su posición anterior.

5.6. Resumen de los resultados obtenidos

En resumen, los resultados obtenidos son los siguientes:

- Se produce supresión perceptual en la condición bimodal en la que está presente el sonido que afecta tanto a la *detección* como al *reconocimiento* de los estímulos visuales. Esta supresión se puede ver en las Figuras 5.1 y 5.3.
- El factor posición del estímulo visual no afecta ni a la detección ni al reconocimiento, como se demuestra en el experimento previo (Anexo A).
- El factor de forma del estímulo visual afecta al reconocimiento de los estímulos visuales. Sin embargo, cuando se tienen en cuenta únicamente estímulos de la condición *condBi* su

influencia deja de ser significativa.

- La presencia del sonido afecta de forma significativa tanto a la detección como al reconocimiento de los estímulos visuales como se puede observar en las Figuras 5.1 y 5.3.
- La localización de la fuente del sonido no influye significativamente ni en la detección ni en el reconocimiento de los estímulos visuales, pese a que *se observa un efecto de disminución de ambos cuando el sonido se presenta directamente desde detrás.*
- Todos los tipos de sonido utilizados influyen tanto en la detección como en el reconocimiento de los estímulos. Es interesante destacar que *ningún usuario ha sido capaz de reconocer correctamente los estímulos visuales asociados al ruido rosa.*
- El análisis cualitativo de los datos de *eyetracker* parece sugerir que uno de los mecanismos del efecto de supresión visual encontrado es la supresión sacádica.

6. Atención en VR

Como se ha explicado en la Sección 3.1, la supresión perceptual y el control de la atención están intrínsecamente relacionados. En la Sección 4.1 se ha demostrado la existencia de un fenómeno de supresión visual intermodal desencadenado por estímulos sonoros. Esta supresión visual se puede integrar en técnicas existentes del control de la atención, como se explica a lo largo de esta Sección.

Se ha demostrado que ciertos comportamientos bien conocidos de la percepción humana se mantienen al introducir a un usuario en un VE [35] mediante replicación de experimentos. Sin embargo, debido a la diferencia inherente entre los medios tradicionales y la VR, es necesario revisar las afirmaciones aceptadas sobre cómo funciona la atención humana en este nuevo medio. Existen cuerpos de trabajo que se dedican a esta labor, como el de Sitzmann et al. [29]. En este trabajo se realiza un estudio en el que se obtiene la medida de saliencia de varias escenas inmersivas a partir de datos obtenidos de varios usuarios que exploran dichas escenas mediante un *eyetracker* montado en las gafas de realidad virtual (HMD por sus siglas en inglés). Una de las conclusiones que obtienen en este trabajo es que existe un sesgo ecuatorial de la atención. Es decir, la gente le presta más atención a lo que se encuentra a lo largo del ecuador de una escena dentro de un entorno virtual (VE). Esto incluye por ejemplo la línea del horizonte, las paredes de una habitación, los objetos que se encuentran a una altura similar a la del observador, etc.

Otro estudio relacionado con la atención en realidad virtual es el de Serrano et al. [30]. En su trabajo, se estudia cómo las ediciones de vídeo aplicadas a la cinematografía tradicional afectan a la experiencia del usuario y a su sensación de continuidad de la historia cuando está viendo una escena en VR. Durante este estudio, proponen una serie de métricas que sirven para medir la atención de un usuario en VR a partir de los datos obtenidos con un *eyetracker* de su exploración de las escenas. También describen una serie de guías para generar nuevo contenido en función de la reacción que se espera suscitar en el espectador. En su trabajo, descubren que distintas técnicas cinematográficas tienen efectos sobre la atención del espectador que hay que tener en cuenta en VR. Otros trabajos se centran de forma más específica en cómo ciertos aspectos de una escena de VR influyen en la atención del usuario. En concreto, en un trabajo pendiente de publicar en el que ha participado la autora de este trabajo, se estudia cómo se perciben distintos materiales en un VE en función del sonido que emiten al ser golpeados por una baqueta de madera.

6.1. Control de la atención en VR

Dirigir la atención de una persona sin que esta sea consciente de ello no es un problema trivial. Cuando se trata de dirigir la atención de un usuario en VR, el reto es aún mayor. Como se ha explicado en la Sección 1.1, el usuario tiene el control total de la cámara en el VE. Bloquear su movimiento natural es una técnica no recomendada, ya que una diferencia entre los movimientos naturales del usuario y los virtuales provoca malestar general, mareos y náuseas entre otros síntomas [83]. Una posible solución es permitir al usuario moverse libremente, pero mostrar contenido únicamente en una zona del espacio a la que el usuario acabará mirando. Sin embargo, este tipo de enfoque malgasta el potencial del entorno inmersivo de 360 grados que permite la tecnología de VR.

Existen otras técnicas que tratan de dirigir la atención del usuario de forma no sutil. En el trabajo de Renner [84] se estudia el uso de flechas, dianas, esferas y túneles para dirigir la atención de los usuarios en tareas de búsqueda en VR. Gruenefeld et al. utilizan puntos de colores para guiar al usuario hacia objetos que se encuentran fuera de su FOV [85]. Incluso existen trabajos que se sirven de distintas modalidades (como la táctil [86]) junto con estímulos visuales para guiar la atención visual. El problema en general de estas técnicas es el intrusismo de los estímulos visuales que presentan. Pese a su efectividad, mostrar estímulos ajenos a la escena puede afectar negativamente a la sensación de inmersión del usuario y a su experiencia. Por ello, en este trabajo se busca una forma de poder dirigir la atención visual del usuario que no presente estímulos visuales adicionales en la escena.

Una alternativa actual es el uso de técnicas de redirección sutil o *subtle gaze direction* [53] que se basan en la utilización de estímulos visuales en la periferia del FOV del usuario. Estos estímulos, correctamente presentados, son prácticamente indetectables. Su funcionamiento se basa en que un cambio del contenido visual en la periferia del FOV desencadena una sacada (un movimiento rápido del ojo durante el cual no se es consciente de los cambios en el entorno hasta cierto punto [87]) cuyo propósito puede ser doble. Por un lado, se puede dirigir la atención hacia el punto de aterrizaje de la sacada para que el usuario siga un *camino visual* mediante sucesivas sacadas hasta llegar a la zona que se le quiere mostrar. Por otro lado, se puede aprovechar el tiempo que dura la sacada para hacer ligeros cambios en el entorno hasta que la zona de interés aparece dentro del FOV del usuario, gracias a la supresión perceptual que ocurre durante la misma. En la Sección 3.1 se discuten más a fondo las desventajas de estos métodos y por qué se ha optado por una propuesta diferente. Sin embargo, la idea de aprovecharse de esta supresión perceptual para conseguir dirigir la atención del usuario en VR sigue siendo la propuesta central de este trabajo.

En resumen, debido a la libertad adicional que el usuario posee en los entornos de VR se vuelve más necesario que nunca ser capaces de dirigir su atención para que no pase por alto ningún contenido interesante. Este proceso o técnica de redirección debe llevarse a cabo de forma sutil, para que no se vea afectada negativamente la sensación de inmersión del usuario y además la técnica utilizada debería ser transferible a un HMD de prestaciones comerciales de forma prácticamente inmediata. Debido a la subjetividad inherente a la percepción humana y a las restricciones de *hardware* establecidas, este es un desafío que todavía no ha sido resuelto. En este trabajo se propone utilizar un efecto de supresión visual producido por un estímulo sonoro como desencadenante para una técnica de control de la atención. En concreto, con el experimento

descrito en la Sección 4.1 se ha obtenido un nuevo método para desencadenar supresión visual en VR.

Propuesta de técnica de control de la atención. Durante este trabajo se ha obtenido una nueva forma de desencadenar la supresión visual en VR mediante un efecto intermodal, en concreto sonoro. La especificación formal y comprobación de una técnica de control de la atención es un trabajo que queda fuera de la envergadura de este trabajo, pero que se desarrollará en un futuro como parte de la tesis de la autora. Sin embargo, a continuación se esboza una técnica de control de la atención que utiliza dicho efecto:

- Se introduce al usuario en una escena de mayor o menor complejidad, idealmente con un sonido ambiente continuo que disimule en cierta medida los estímulos sonoros que se van a utilizar.
- Durante la tarea que el usuario esté realizando (ya sea búsqueda, exploración o cualquier otra) se reproducirán los sonidos que desencadenan la supresión perceptual.
- Gracias al experimento realizado en este trabajo, sabemos que existe cierta probabilidad de que pasados 100ms del inicio del sonido haya una supresión de por lo menos 24ms.
- El tiempo de supresión visual se aprovecha para modificar el entorno del usuario, ya sea moviendo el mismo respecto al punto de vista del usuario [3], o realizando pequeños cambios en el entorno como la aparición o desaparición de ciertos objetos.
- Este fenómeno puede repetirse las veces que sea necesario, puesto que no se ha encontrado ningún signo de habituación en el experimento realizado (hasta las 18 repeticiones que se han probado).

Este fenómeno puede aplicarse a cualquier técnica que utilice la supresión visual de forma modular, sin necesidad de modificar ningún otro aspecto.

7. Aplicaciones a la medicina

Como se ha explicado en el Capítulo 3 la VR ya se utiliza en medicina. En concreto, uno de los tratamientos en los que la atención del paciente es decisiva es el tratamiento de exposición para fobias. Una parte importante del tratamiento de cualquier fobia es la exposición al estímulo temido hasta que este deja de producir temor. En la mayoría de los casos, el temor se mantiene por el refuerzo negativo que resulta de evitar el estímulo temido [88]. Una de las componentes características de una fobia es la conducta de evitación y escape, que incluye la distracción del desencadenante del miedo mediante la fijación de la atención en otros estímulos.

La realidad virtual ya se utiliza en el tratamiento de exposición con resultados positivos [62] para el paciente ya que presenta varias ventajas sobre un tratamiento de exposición tradicional. Entre ellas, el terapeuta posee un mayor control sobre el nivel de exposición y el paciente se encuentra en un entorno que no supone un peligro real para él y que le permite afrontar su miedo de forma gradual y segura.

Sin embargo, dada la fuerte componente que supone la atención del paciente en la fobia, se puede suponer que una técnica del control de la atención sería capaz de influir hasta cierto punto en la rapidez de su recuperación. En esta Sección se describe una propuesta de tratamiento de exposición utilizando control de la atención en VR mediante supresión visual desencadenada por estímulos sonoros. Este tratamiento de exposición no ha sido probado todavía con ningún paciente que sufra de aracnofobia, por lo que un caso de uso real permanece como trabajo futuro.

7.1. Propuesta de tratamiento de exposición para aracnofobia

Una de las ventajas de este tipo de tratamientos, es que pueden servir de apoyo al tratamiento realizado por el terapeuta [89]. Fuera de la consulta, el paciente puede decidir realizar una o varias sesiones de exposición con la aplicación propuesta en un entorno seguro como su casa sabiendo que no correrá ningún peligro real. Este tratamiento podría llevarse a cabo con distintos grados de implicación por parte del terapeuta, desde sesiones semanales de seguimiento a no tener ningún tipo de contacto con el paciente durante el experimento. Los aspectos a seguir durante este tipo de terapia se enumeran a continuación.

- Inicialmente, el paciente rellena un cuestionario sobre su miedo a las arañas [90]. Además, contesta a preguntas sobre su estado de ánimo y el grado de miedo y evitación que sentiría hacia una araña pequeña, mediana y grande. Por último, el paciente indica el grado de impedimento que le supone su fobia (miedo paralizante, huida, evasión, etc.) y cuánto

espera mejorar con el tratamiento [91].

- Si el terapeuta está presente, este realizará una evaluación clínica del miedo del paciente a una araña real.
- **Procedimiento:** Los pacientes se colocan el casco y aparecen en una habitación virtual. En las primeras etapas del tratamiento, o si su nivel de ansiedad o miedo es muy alto, se encontrarán con un entorno amigable como el de la Figura 7.2 (Arriba). Cuando disminuye el miedo a lo largo de las sesiones y el paciente se considera preparado, se le introduce en un entorno hostil (Figura 7.2, abajo). Este cambio en el entorno busca romper con la monotonía y suscitar una respuesta emocional en el paciente, para evitar la sensación de control sobre el mismo. Ambos escenarios presentan una radio o televisión.
  - Primero se permite al usuario explorar la habitación durante dos minutos sin que pase nada, para que se acostumbre al nuevo entorno.
  - Pasado este tiempo, la araña (Figura 7.1) aparece en el campo de visión. Idealmente, se ve cómo sale andando desde detrás de alguno de los muebles de las estancias para darle una sensación de realismo mayor. Al principio, la araña aparece lejos del paciente, pero dentro de su campo de visión.
  - Si el paciente evita mirar a la araña durante más de 20 segundos, se aleja de ella o directamente mira en otra dirección se inicia el sistema de redireccionamiento con supresión visual. La radio o televisión del entorno servirán como sonido ambiente de la escena. Sobre ese sonido se reproducirá un estímulo adicional que cause la supresión visual. Teniendo en cuenta que los seis sonidos utilizados en el experimento de la Sección 4.1 son capaces de desencadenar la supresión visual (Sección 5.3), se elige utilizar los tres sonidos de ruido camuflados como interferencias del sonido ambiente.
  - En este caso, cada vez que el sonido produzca una supresión perceptual (aproximadamente 100ms después del inicio de la reproducción) se modificará el entorno alrededor del paciente (de la misma forma que Sun et al. [3] modifica la escena sin que el usuario sea consciente de ello) para mantener a la araña siempre dentro de su campo de visión.
  - Cuando el paciente se acostumbre a tener a la araña en su campo de visión, la araña se acercará a él paulatinamente. Cada vez que la araña se acerque se repetirá el proceso de supresión perceptual si es necesario. Si el paciente cuenta con alguien que le ayude, incluso puede utilizar un peluche similar a la araña para tocarlo cuando la araña esté a su lado.
- El paciente debe intentar superar sus límites, pero sabiendo siempre que puede detener el experimento si este resulta ser demasiado.
- Después del tratamiento el paciente vuelve a rellenar los cuestionarios sobre su miedo a las arañas y ansiedad. Si no se realizan más sesiones, estos cuestionarios pueden volver a rellenarse como seguimiento en distintos intervalos de tiempo que van desde una semana después del procedimiento hasta un año después [91].

Figura 7.1: Modelo escogido para diseñar una escena de exposición para el tratamiento de aracnofobia.

Figura 7.2: Entornos escogidos para la escena de exposición del tratamiento de aracnofobia. Estos entornos han sido escogidos según el nivel de respuesta emocional que pueden desencadenar. Arriba: entorno amigable, para los primeros pasos de la exposición. Abajo: entorno frío para aumentar el nivel de estrés o desasosiego del paciente.

8. Conclusiones y trabajo futuro

En resúmen, los resultados obtenidos del experimento realizado durante este trabajo son los siguientes:

- Se produce supresión perceptual en la condición bimodal debido a la presencia del sonido.
- El factor de forma del estímulo pierde su influencia significativa en el reconocimiento de los estímulos cuando se analiza únicamente la condición bimodal.
- La presencia del sonido afecta de forma significativa tanto a la detección como al reconocimiento de los estímulos visuales independientemente del tipo de sonido usado o su localización.
- Una de las posibles causas del efecto observado es la supresión sacádica.

En este proyecto se ha demostrado, hasta donde sabemos, por primera vez, que el fenómeno de supresión visual causada por un estímulo auditivo tiene lugar en un entorno de VR con estímulos complejos. Cuando se reproduce un sonido fuera del campo de visión de un usuario en VR se desencadena un efecto de supresión perceptual que afecta a la visión, por lo que el usuario no es consciente de los cambios en los estímulos visuales que tiene delante de sí. Esto sólo se había comprobado en un entorno controlado y en medio tradicional, en el trabajo de Hidaka et al. [76], con estímulos muy concretos (parches de Gabor) y teniendo en cuenta únicamente el efecto sobre el reconocimiento o discriminación del estímulo visual. Este descubrimiento es útil como método adicional para generar una supresión visual, ya utilizada en trabajos existentes tanto en VR como fuera de ella. De hecho, el resultado inmediato de este trabajo será una publicación en la que ya se está trabajando.

En cuanto al efecto de supresión, éste ha resultado ser robusto a variaciones de factores. Los seis tipos de sonidos probados en tres localizaciones diferentes han sido capaces de desencadenar la supresión sobre las cinco formas diferentes que se han explorado, y todos los participantes del experimento descrito en la Sección 4.1 se han visto afectados por el mismo, independientemente de su sexo, edad o experiencia previa con VR. Este efecto no sólo influye en la detección del estímulo visual, si no también en su reconocimiento, ya que además del bajo porcentaje de estímulos *detectados* en la condición bimodal (18.25 %), el porcentaje de estímulos correctamente *reconocidos* es aún menor (5 %).

Aún así, quedan fronteras abiertas, y muchas oportunidades de trabajo futuro interesantes. Por ejemplo, un análisis en profundidad del espacio de los sonidos. La idea inicial de este trabajo

era encontrar un estímulo no intrusivo que desencadenase supresión perceptual. ¿Es posible generar este mismo efecto con sonidos más sutiles o con frecuencias que el oído humano no percibe de forma consciente? Durante este trabajo se han realizado pruebas y evaluaciones conservadoras pero, ¿qué pasaría si los estímulos visuales y auditivos se integrasen semánticamente con la escena?, ¿aumentaría el efecto de supresión del sonido?

Otra línea de trabajo futuro interesante es un estudio formal y robusto de los datos de *eyetracker* de los usuarios. De los 54 estímulos presentados, 18 no poseen componente visual, por lo que no se ha obtenido una respuesta directa de los participantes respecto a ellos. ¿Qué ocurre cuando se reproducen estos sonidos? Por otro lado, en la sección de análisis se ha sugerido que el efecto puede deberse a la supresión sacádica, pero aún se desconoce si ese es el único mecanismo que influye en la supresión visual o si es un comportamiento común a todos los usuarios.

Como se comenta en la Sección 4.1, ningún participante tenía problemas auditivos. ¿En qué medida se mantendrá este efecto en personas que están perdiendo la sensibilidad ante las frecuencias más agudas debido a la edad, o en personas que sólo oyen correctamente por uno de los oídos?

Saber que la atención visual puede verse suprimida por un estímulo de otra modalidad reabre la pregunta de cómo de relacionadas están la atención visual y la auditiva. Una línea de trabajo futuro interesante es el estudio de modelos de atención multimodal en VR que ayuden a comprender mejor las interacciones de los distintos sistemas sensoriales a la hora de formar una percepción del entorno virtual que rodea al usuario. Existen modelos en medios tradicionales, como el de Min et al. [41] que tienen en cuenta esta interacción. Trasladarlo a VR, añadiendo los efectos de supresión, permitirá obtener una mejor idea de cómo se relacionan los usuarios con el entorno.

Finalmente, se ha esbozado en el Capítulo 7 un posible tratamiento de exposición para la aracnofobia. Para saber realmente si este tratamiento mejora en algo con respecto a los tratamientos actuales debe realizarse un estudio formal con un grupo de control, otro que se someta a terapia convencional y un último que siga el método propuesto en este trabajo. Este estudio debería llevarse a cabo junto con uno o varios profesionales de la salud que pudiesen aportar su punto de vista en cuanto a mejoras y necesidades para sus pacientes.

Personalmente, este trabajo ha supuesto una forma de expandir los conocimientos asentados en el máster. Este trabajo servirá como primera piedra de mi tesis de doctorado en un tema que cada vez me resulta más apasionante. Explorar problemas aún sin respuesta supone nuevos e interesantes retos que afrontar continuamente. Pese a que el trabajo a veces me ha puesto a prueba, considero que he aprendido muchas cosas que antes desconocía y lejos de verlo como un punto y final lo considero como un camino abierto a un gran número de posibilidades para los siguientes pasos de mi vida académica. De nuevo gracias a mis directoras y al grupo de investigación en el que se ha desarrollado este TFM he podido comenzar un programa de doctorado que tampoco será el final de mi vida académica.

Bibliografía

[1] Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. Can simulator sickness be avoided? a review on temporal aspects of simulator sickness. *Frontiers in Psychology*, 9:2132, 2018.

[2] Eike Langbehn and Frank Steinicke. Redirected walking in virtual reality, 2018.

[3] Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan McGuire, David Luebke, and Arie Kaufman. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Transactions on Graphics (TOG)*, 37(4):67, 2018.

[4] Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics (TOG)*, 36(4):50, 2017.

[5] Graham Hankinson. The brand images of tourism destinations: a study of the saliency of organic images. *Journal of Product & Brand Management*, 13(1):6:14, 2004.

[6] Claudio Cifarelli. Content based image selection for automatic photo album generation, 2013. US Patent 8,571,331.

[7] Chenlei Guo and Liming Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Processing*, 19(1):185–198, 2010.

[8] Nicholas J Butko, Lingyun Zhang, Garrison W Cottrell, and Javier R Movellan. Visual saliency model for robot cameras. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 2398–2403. IEEE, 2008.

[9] Luca Marchesotti, Claudio Cifarelli, and Gabriela Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2232–2239. IEEE, 2009.

[10] Dwarikanath Mahapatra and Ying Sun. Rigid registration of renal perfusion images using a neurobiology-based visual saliency model. *Journal on Image and Video Processing*, 2010:4, 2010.

[11] Hajime Hata, Hideki Koike, and Yoichi Sato. Visual guidance with unnoticed blur effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 28–35. ACM, 2016.

[12] Ann McNamara, Reynold Bailey, and Cindy Grimm. Search task performance using subtle gaze direction with the presence of distractions. *ACM Transactions on Applied Perception (TAP)*, 6:17, 2009.

[13] Pawel Kiper, Andrzej Szczudlik, Michela Agostini, Jozef Opara, Roman Nowobilski, Laura Ventura, Paolo Tonin, and Andrea Turolla. Virtual reality for upper limb rehabilitation in subacute and chronic stroke: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*, 99(5):834 – 842.e4, 2018.

[14] Kumar Raghav Gujjar, Arjen van Wijk, Ratika Sharma, and Ad de Jongh. Virtual reality exposure therapy for the treatment of dental phobia: A controlled feasibility study. *Behavioural and Cognitive Psychotherapy*, 46(3), 2018.

[15] Sistema nervioso sensorial. `https://askabiologist.asu.edu/que_hace_tu_cerebro`. Accedido por última vez el 04 de noviembre de 2018.

[16] Sistema visual. `https://es.wikipedia.org/wiki/Ojo_humano#/media/File:Eyesection-es.svg`. Accedido por última vez el 04 de noviembre de 2018.

[17] Sistema auditivo. `http://liceu.uab.es/~joaquim/phonetics/fon_percept/audicio/audicion.html`. Accedido por última vez el 04 de noviembre de 2018.

[18] James W. Lewis, Michael S. Beauchamp, and Edgar A. DeYoe. A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, 10(9):873–888, 2000.

[19] David A Bulkin and Jennifer M Groh. Seeing sounds: visual and auditory interactions in the brain. *Current opinion in neurobiology*, 16(4):415–419, 2006.

[20] Ilana B. Witten and Eric I. Knudsen. Why seeing is believing: Merging auditory and visual worlds. *Neuron*, 48(3):489 – 496, 2005.

[21] Charles Spence, Jae Lee, and Nathan Van der Stoep. Responding to sounds from unseen locations: crossmodal attentional orienting in response to sounds presented from the rear. *European Journal of Neuroscience*, 2017.

[22] Petra Vetter, Fraser W Smith, and Lars Muckli. Decoding sound and imagery content in early visual cortex. *Current Biology*, 24(11):1256–1262, 2014.

[23] Mark A Eckert, Nirav V Kamdar, Catherine E Chang, Christian F Beckmann, Michael D Greicius, and Vinod Menon. A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fmri connectivity analysis. *Human brain mapping*, 29(7):848–857, 2008.

[24] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Illusions: What you see is what you hear. *Nature*, 408(6814):788, 2000.

[25] Jeremy N Bailenson, Nick Yee, Jim Blascovich, Andrew C Beall, Nicole Lundblad, and Michael Jin. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, 17(1):102–141, 2008.

[26] Wolfgang Müller-Wittig. Virtual reality in medicine. In *Springer Handbook of Medical Technology*, pages 1167–1186. Springer, 2011.

[27] Gregory I Gewickey, Lewis S Ostrover, Michael Smith, and Michael Zink. Immersive virtual reality production and playback for storytelling content, October 23 2018. US Patent App. 10/109,320.

[28] EC Hamilton, DJ Scott, JB Fleming, RV Rege, R Laycock, PC Bergen, ST Tesfay, and DB Jones. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy and Other Interventional Techniques*, 16(3):406–411, 2002.

[29] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE transactions on visualization and computer graphics*, 24(4):1633–1642, 2018.

[30] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (TOG)*, 36(4):47, 2017.

[31] Randolph Blake. A neural theory of binocular rivalry. *Psychological review*, 96:145, 1989.

[32] Yoram S Bonneh, Alexander Cooperman, and Dov Sagi. Motion-induced blindness in normal observers. *Nature*, 411:798, 2001.

[33] Naotsugu Tsuchiya and Christof Koch. Continuous flash suppression reduces negative afterimages. *Nature neuroscience*, 8:1096, 2005.

[34] Ethel Matin. Saccadic suppression: a review and an analysis. *Psychological bulletin*, 81(12):899, 1974.

[35] Marcos Allue, Ana Serrano, Manuel G. Bedia, and Belen Masia. Crossmodal Perception in Immersive Environments. In *Spanish Computer Graphics Conference (CEIG)*. The Eurographics Association, 2016.

[36] Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F Welch, and Gerd Bruder. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics (TOG)*, 37(4):66, 2018.

[37] Benjamin Bolte and Markus Lappe. Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. *IEEE transactions on visualization and computer graphics*, 21(4):545–552, 2015.

[38] Brian J White, David J Berg, Janis Y Kan, Robert A Marino, Laurent Itti, and Douglas P Munoz. Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature communications*, 8:14263, 2017.

[39] Peter M Forster and Ernest Govier. Discrimination without awareness? *The Quarterly Journal of Experimental Psychology*, 30(2):289–295, 1978.

[40] Jeremy M Wolfe, Melissa L-H Võ, Karla K Evans, and Michelle R Greene. Visual search in scenes involves selective and nonselective pathways. *Trends in cognitive sciences*, 15(2):77–84, 2011.

[41] Xiongkuo Min, Guangtao Zhai, Ke Gu, and Xiaokang Yang. Fixation prediction through multimodal analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 13(1):6, 2017.

[42] Jeremy M Wolfe and W Gray. Guided search 4.0. *Integrated models of cognitive systems*, pages 99–119, 2007.

[43] Jochen Ditterich, Thomas Eggert, and Andreas Straube. The role of the attention focus in the visual information processing underlying saccadic adaptation. *Vision research*, 40(9):1125–1134, 2000.

[44] Anna C Nobre, DR Gitelman, EC Dias, and MM Mesulam. Covert visual spatial orienting and saccades: overlapping neural systems. *Neuroimage*, 11(3):210–216, 2000.

[45] Charles Spence and Jon Driver. Audiovisual links in exogenous covert spatial orienting. *Perception & psychophysics*, 59(1):1–22, 1997.

[46] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pages 1597–1604. IEEE, 2009.

[47] Joseph H Goldberg and Jonathan I Helfman. Visual scanpath representation. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 203–210. ACM, 2010.

[48] Nicholas Waldin, Manuela Waldner, and Ivan Viola. Flicker observer effect: Guiding attention through high frequency flicker in images. In *Computer Graphics Forum*, volume 36, pages 467–476. Wiley Online Library, 2017.

[49] Irwin Pollack and James M Pickett. Cocktail party effect. *The Journal of the Acoustical Society of America*, 29(11):1262–1262, 1957.

[50] Ziad M Hafed and James J Clark. Microsaccades as an overt measure of covert attention shifts. *Vision research*, 42(22):2533–2545, 2002.

[51] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.

[52] Sean Andrist, Michael Gleicher, and Bilge Mutlu. Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2571–2582. ACM, 2017.

[53] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. Subtle gaze direction. *ACM Transactions on Graphics (TOG)*, 28(4):100, 2009.

[54] Frank Bauer, Samuel W Cheadle, Andrew Parton, Hermann J Müller, and Marius Usher. Gamma flicker triggers attentional selection without awareness. *Proceedings of the National Academy of Sciences*, 106(5):1666–1671, 2009.

[55] Kuno Kurzhals, Markus Höferlin, and Daniel Weiskopf. Evaluation of attention-guiding video visualization. In *Computer Graphics Forum*, volume 32, pages 51–60. Wiley Online Library, 2013.

[56] Margarita Vinnikov, Robert S Allison, and Suzette Fernandes. Gaze-contingent auditory displays for improved spatial attention in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(3):19, 2017.

[57] Brooke E Wooley and David S March. Exploring the influence of audio in directing visual attention during dynamic content. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 187–190. ACM, 2014.

[58] Ali Pourmand, Steven Davis, Danny Lee, Scott Barber, and Neal Sikka. Emerging utility of virtual reality as a multidisciplinary tool in clinical medicine. *Games for health journal*, 6(5):263–270, 2017.

[59] JL Mosso Vasquez, Verónica Lara Vaca, Brenda K Wiederhold, I Miller, and Mark D Wiederhold. Virtual reality pain distraction during gynecological surgery: A report of 44 cases. *Surgical Research Updates*, 2017.

[60] Mar Rus-Calafell, Philippa Garety, Elinor Sason, Thomas JK Craig, and Lucia R Valmaggia. Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. *Psychological medicine*, 48(3):362–391, 2018.

[61] Christoph Guger, Brendan Allison, Fan Cao, and Guenter Edlinger. A brain-computer interface for motor rehabilitation with functional electrical stimulation and virtual reality. *Archives of Physical Medicine and Rehabilitation*, 98(10):e24, 2017.

[62] Per Carlbring. Single-session gamified virtual reality exposure therapy for spider phobia vs. traditional exposure therapy: A randomized-controlled non-inferiority trial with 12-month follow-up. In *Anxiety and Depression Association of America Conference, San Francisco, USA, 6-9 April 2017.*, 2017.

[63] YQ Zhou, C Li, CY Shui, YC Cai, RH Sun, DF Zeng, W Wang, QL Li, L Huang, J Tu, and J Jiang. [application of virtual reality in surgical treatment of complex head and neck carcinoma]. *Chinese journal of otorhinolaryngology head and neck surgery*, 53(1):49:52, January 2018.

[64] Filip Górski, Paweł Buń, Radosław Wichniarek, Przemysław Zawadzki, and Adam Hamrol. Effective design of educational virtual reality applications for medicine using knowledge-engineering techniques. *Eurasia Journal of Mathematics, Science & Technology Education*, 13(2), 2017.

[65] D. Duncan, B. Newman, A. Saslow, E. Wanserski, T. Ard, R. Essex, and A. Toga. Vrain: Virtual reality assisted intervention for neuroimaging. In *2017 IEEE Virtual Reality (VR)*, 2017.

[66] Timur Kuzhagaliyev, Neil T Clancy, Mirek Janatka, Kevin Tchaka, Francisco Vasconcelos, Matthew J Clarkson, Kurinchi Gurusamy, David J Hawkes, Brian Davidson, and Danail Stoyanov. Augmented reality needle ablation guidance tool for irreversible electroporation in the pancreas. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions,*

and Modeling, volume 10576, page 1057613. International Society for Optics and Photonics, 2018.

[67] Michael Connolly, Johnathan Seligman, Andrew Kastenmeier, Matthew Goldblatt, and Jon C. Gould. Validation of a virtual reality-based robotic surgical skills curriculum. *Surgical Endoscopy*, 28(5):1691–1694, May 2014.

[68] Jin Guo, Shuxiang Guo, Takashi Tamiya, Hideyuki Hirata, and Hidenori Ishihara. A virtual reality-based method of decreasing transmission time of visual feedback for a tele-operative robotic catheter operating system. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 12(1):32–45, 2016.

[69] Jillian L McGrath, Jeffrey M Taekman, Parvati Dev, Douglas R Danforth, Deepika Mohan, Nicholas Kman, Amanda Crichlow, William F Bond, Shiela Riker, AJ Lemheney, et al. Using virtual reality simulation environments to assess competence for emergency medicine learners. *Academic Emergency Medicine*, 25(2):186–195, 2018.

[70] Nicholas Raison, Kamran Ahmed, Nicola Fossati, Nicolò Buffi, Alexandre Mottrie, Prokar Dasgupta, and Henk Van Der Poel. Competency based training in robotic surgery: benchmark scores for virtual reality robotic simulation. *Bju international*, 119(5):804–811, 2017.

[71] Realidad virtual para el cuidado del paciente. `https://www.limbix.com/`. Accedido por última vez el 12 de noviembre de 2018.

[72] Noticia: uso de vr en hospital para aumentar la tolerancia al dolor. `https://goo.gl/Y22aPh`. Accedido por última vez el 12 de noviembre de 2018.

[73] John Ross, M Concetta Morrone, Michael E Goldberg, and David C Burr. Changes in visual perception at the time of saccades. *Trends in neurosciences*, 24(2):113–121, 2001.

[74] Maarten A Frens, A John Van Opstal, and Robert F Van der Willigen. Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, 57(6):802–816, 1995.

[75] Patrycja Delong, Máté Aller, Anette S Giani, Tim Rohe, Verena Conrad, Masataka Watanabe, and Uta Noppeney. Invisible flashes alter perceived sound location. *Scientific reports*, 8(1):12376, 2018.

[76] Souta Hidaka and Masakazu Ide. Sound can suppress visual perception. *Scientific reports*, 5:10483, 2015.

[77] Pierre Salamé and Alan Baddeley. Noise, unattended speech and short-term memory. *Ergonomics*, 30(8):1185–1194, 1987.

[78] Sonido ambiente *open Source* de un parque. `https://freesound.org/home/login/?next=/people/klankbeeld/sounds/387204/`. Accedido por última vez el 12 de noviembre de 2018.

[79] Rtve a la carta: descarga de *podcasts* de rne5. `http://www.rtve.es/alacarta/rne/radio-5/`. Accedido por última vez el 12 de noviembre de 2018.

[80] Unity 3d living room - by barking dog. https://assetstore.unity.com/packages/3d/environments/urban/3d-living-room-62120. Accedido por última vez el 04 de noviembre de 2018.

[81] BD Corneil, M Van Wanrooij, DP Munoz, and AJ Van Opstal. Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology*, 88:438:454, 2002.

[82] John L Sibert, Mehmet Gokturk, and Robert A Lavine. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, page 101:107. ACM, 2000.

[83] Charles M Oman. Sensory motor conflict theory for motion sickness. *Letter to Reason*, 2018.

[84] Patrick Renner. Prompting techniques for guidance and action assistance using augmented-reality smart-glasses. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 820–822. IEEE, 2018.

[85] Uwe Gruenefeld, Andreas Löcken, Yvonne Brueck, Susanne Boll, and Wilko Heuten. Where to look: Exploring peripheral cues for shifting attention to spatially distributed out-of-view objects. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 221–228. ACM, 2018.

[86] Tim Claudius Stratmann, Andreas Loecken, Uwe Gruenefeld, Wilko Heuten, and Susanne Boll. Exploring vibrotactile and peripheral cues for spatial attention guidance. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, page 9. ACM, 2018.

[87] Wikipedia: sacadas. https://en.wikipedia.org/wiki/Saccade. Accedido por última vez el 13 de noviembre de 2018.

[88] Martin M Antony and David H Barlow. Specific phobia. *International handbook of cognitive and behavioural treatments for psychological disorders*, page 1:22, 1998.

[89] A Ghosh, Isaac M Marks, and AC Carr. Therapist contact and outcome of self-exposure treatment for phobias: A controlled study. *The British Journal of Psychiatry*, 152(2):234–238, 1988.

[90] Jeff Szymanski and William O'Donohue. Fear of spiders questionnaire. *Journal of behavior therapy and experimental psychiatry*, 26:31:34, 1995.

[91] Lars-Göran Öst, Paul M Salkovskis, and Kerstin Hellström. One-session therapist-directed exposure vs. self-exposure in the treatment of spider phobia. *Behavior Therapy*, 22(3):407–422, 1991.

Anexo A. Experimento previo

A lo largo de este anexo se describe el experimento previo realizado para asegurar que los estímulos visuales utilizados en el experimento descrito en la Sección 4.1 son visibles en condiciones idénticas sin la presencia de estímulos auditivos. De esta forma se asegura que el hecho de que el participante no vea los estímulos visuales no se debe a que estos se muestran durante demasiado poco tiempo o a que son indistinguibles del fondo de la escena.

A.1. Descripción del experimento

Participantes. Un total de 7 participantes realizaron el experimento. La media de edad de los participantes era de 27 años con una desviación estándar de 4.6 años. Del total, 1 sujeto era mujer. Todos los sujetos tenían visión normal o corregida y no presentaban problemas auditivos. Los participantes conocían el objetivo del experimento descrito en la Sección 4.1, pero no el de este experimento.

Equipamiento. El *hardware* utilizado durante el experimento ha sido un sistema de RV completo HTC Vive Pro (las gafas de VR (HMD) y dos *trackers* con una superficie calibrada de 4x1.5m por la que el usuario podía moverse libremente durante el experimento). Integrado en el sistema se encuentra un *eyetracker* de Pupil Labs (frecuencia de muestreo = 120Hz). Las especificaciones completas del HMD y del *eyetracker* se pueden consultar en la Sección 2.2. Un único ordenador se utilizaba para el experimento, con un procesador i7-7700 a 3.6GHz, 16GB de RAM y una tarjeta gráfica Nvidia 1060GTX con 6GB de memoria DDR5 dedicados. En cuanto al *software*, todas las escenas fueron programadas en Unity 3D (versión 2018), utilizando los *plug-in* de Pupil Labs para grabar, el *software* de *Steam VR* para integrar el sistema de VR y el *plug-in* de captura *VR-capture* disponible para Unity. El SO utilizado es Windows 10.

Estímulos visuales. Los estímulos visuales (Figura 4.1) consistían en cinco formas simples (círculo, cuadrado, rombo, pentágono y estrella de cinco puntas) con relleno blanco y un borde de un grosor del 5 % del tamaño de la forma de color gris para evitar que el estímulo se pudiera confundir con un fondo blanco. El tamaño de los estímulos es de un grado en el campo visual, y al aparecer, los estímulos visuales permanecen 24ms en el campo de visión del usuario. Tanto el tamaño del estímulo como el tiempo de visualización se han fijado siguiendo el trabajo de Hidaka et al. [76], en el que se demuestra que los estímulos auditivos pueden empeorar la discriminación de estímulos visuales en una pantalla convencional. El estímulo visual puede aparecer en tres localizaciones posibles: *visFront*, *visRight* y *visLeft*. En *visFront*, el estímulo aparece en el centro del campo de visión del usuario. En *visRight* y *visLeft* el estímulo aparece desplazado

cuatro grados hacia la derecha o la izquierda, respectivamente. La razón de utilizar estas tres localizaciones es doble. Por un lado, comprobar si los estímulos se detectan de forma más precisa cuando aparecen en el centro del campo de visión que cuando aparecen desplazados a la izquierda o la derecha. Por otro lado, evitar que el usuario se acostumbre a verlos aparecer siempre en el mismo lugar. Se decide mantener la posición vertical a lo largo del ecuador del campo de visión para facilitar la tarea de detección y reconocimiento, ya que sabemos gracias al trabajo de Sitzmann et al. [29] que es a lo largo de este ecuador donde más se concentra la atención de un usuario en VR.

Estímulos sonoros. En este experimento no existen estímulos auditivos propiamente dichos, pero sí que se presenta un sonido de ambiente igual al del experimento descrito en la Sección 4.1. En concreto, se oye el mismo *podcast* de noticias en español y el sonido ambiente del parque, en las mismas localizaciones. En el siguiente experimento, descrito en la Sección 4.1, este sonido ambiente también estará presente. Utilizarlo en este experimento previo permite obtener una línea base de comparación en las mismas condiciones que se usarán más adelante.

Procedimiento. Los participantes se encuentran en la misma sala virtual que se utilizará en el siguiente experimento (Figura 4.2). Aparecen en el mismo lugar y tienen disponible el mismo espacio para moverse libremente. Al participante se le explica que verá aparecer formas visuales sencillas en su campo visual y que cuando vea una deberá avisar al experimentador. Este pulsa un botón del teclado y aparece una pantalla traslúcida en las cuatro paredes de la habitación con la pregunta *¿Qué ha visto?*. En ese momento el experimentador pregunta al participante cuál es el estímulo que ha visto para registrar su respuesta. El usuario contesta con el nombre de la forma que ha visto si la ha reconocido, o dice que no la ha reconocido si ese es el caso. El experimentador guarda la respuesta del usuario pulsando otro botón del teclado (uno distinto para cada estímulo visual, otro para cuando el usuario ha visto algo pero no ha reconocido la forma) y el experimento continúa. Si el experimentador ha pulsado el primer botón del teclado por error sin que el participante le haya indicado nada, existe otra tecla para indicar esta posibilidad y anular el registro de la respuesta.

Durante el experimento aparecen 50 estímulos visuales aleatorios que el participante debe reconocer. Estos estímulos aparecen en un intervalo aleatorio de entre 5 y 10 segundos y se distribuyen uniformemente en las localizaciones *visFront*, *visRight* y *visLeft*. Los resultados de este experimento se utilizan como línea de base para la condición sólo visual en contraposición con la condición audiovisual *condBi* del experimento de la Sección 4.1. El objetivo primordial de este experimento es evaluar la visibilidad de las formas en las tres localizaciones posibles (*visFront*, *visRight* y *visLeft*). Como información secundaria, también se obtiene una medida de la tasa de discriminación de dichos estímulos visuales. En el primer caso hablamos de *detección* de la presencia de un estímulo visual, y en el segundo de *reconocimiento* del estímulo visto entre varias formas posibles. La hipótesis es que la tasa de detección será superior a la de reconocimiento para este experimento.

Antes de comenzar, el usuario recibe una explicación del procedimiento del experimento y el sistema para comunicar las respuestas al experimentador. El usuario es informado de que si siente cualquier tipo de malestar o mareo debe avisar al experimentador para que pare el experimento. Antes de que el experimentador le ponga el HMD, el participante rellena la primera parte del cuestionario del Anexo B. Una vez puesto el HMD, se realiza una calibración del *eyetracker* en la que el usuario debe mirar fijamente a un punto que se mueve por siete localizaciones fijas de la

pantalla. Después de realizar el experimento, el usuario rellena la segunda parte del cuestionario.

A.2. Resultados

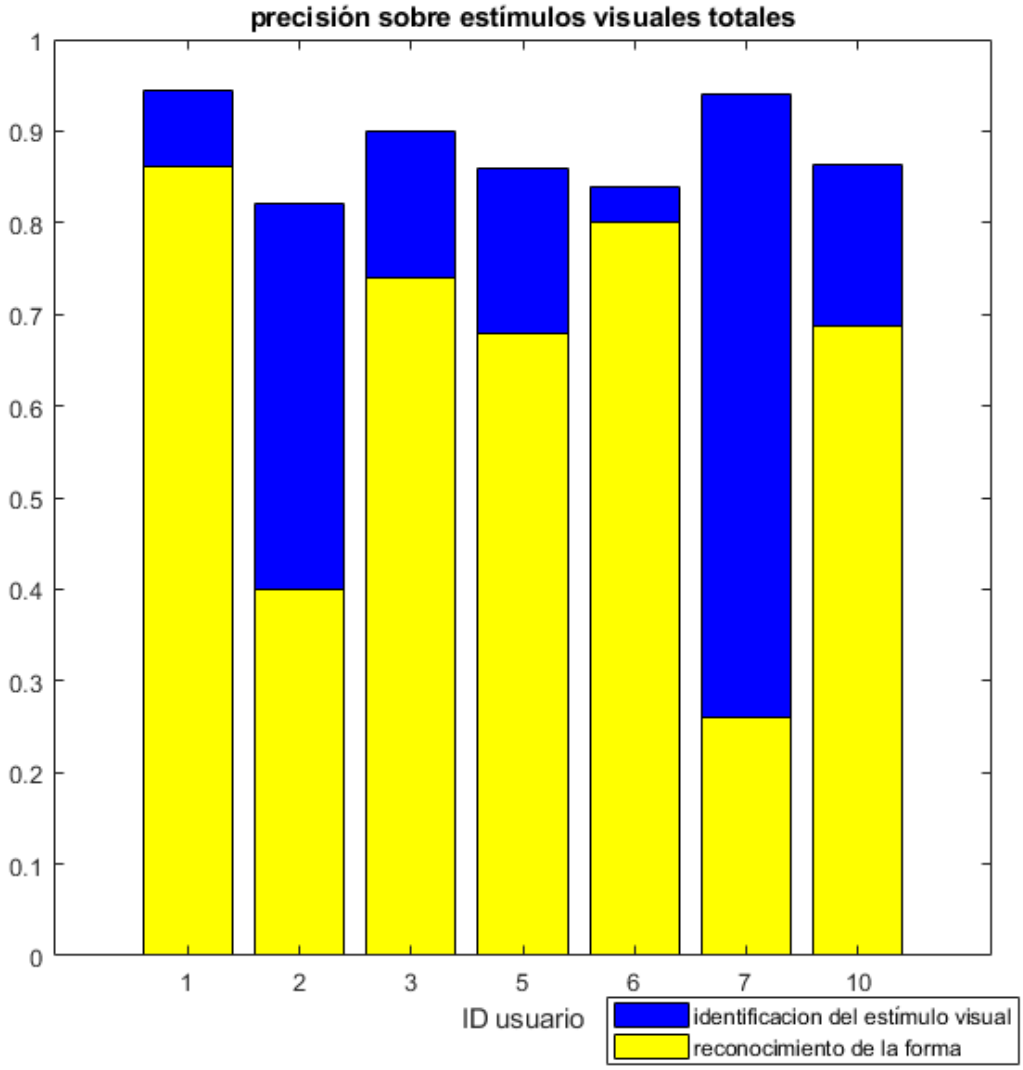


Figura A.1: Resultados para cada uno de los 7 usuarios que llevaron a cabo el test. En azul detección sobre el total de estímulos y en amarillo reconocimiento sobre el total de estímulos para cada usuario.

En media, los participantes detectaron un 88.10 % (±5 % desviación típica) de los estímulos visuales, de los cuales reconocieron correctamente un 71.96 % (±25 %). En la Figura A.1 se puede observar el porcentaje de detección y reconocimiento para cada uno de los participantes del experimento. Como se esperaba, el porcentaje de detección es mayor que el de reconocimiento para todos los participantes.

Al agrupar según las formas mostradas durante el experimento (Figura A.2) se observa

una tasa de aciertos razonable para las cinco formas distintas presentadas en el experimento. Pese a que la tasa de detección es alta en todos los casos, el círculo tiene un porcentaje de reconocimiento ligeramente superior y el pentágono inferior, por lo que es posible que la forma del estímulo afecte al porcentaje de reconocimiento. En cuanto a la posición del estímulo visual (Figura A.3) se observan porcentajes similares para las tres posibles localizaciones. En las Figuras A.4 y A.5 se puede observar un desglose por cada usuario de estos dos factores.

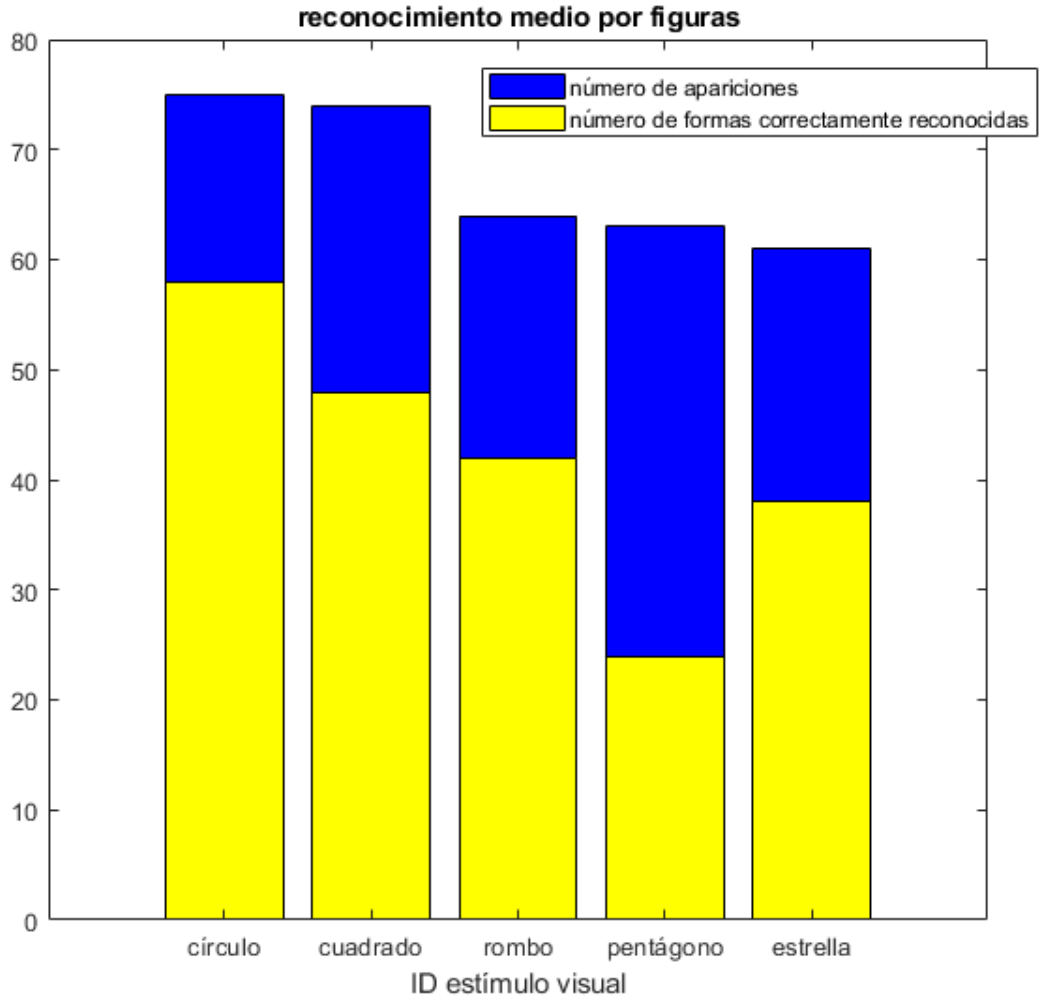


Figura A.2: Detección y reconocimiento medio por forma de estímulo visual. En azul detección sobre el total de estímulos y en amarillo reconocimiento sobre el total para cada forma.

Análisis estadístico. Se ha utilizado un *Generalized linear mixed model* (GLMM) para realizar el análisis estadístico sobre la influencia del factor *forma* y el factor *posición* sobre los porcentajes de detección y reconocimiento de los estímulos visuales presentados en este experimento. La variable respuesta se ha binarizado y es modelada como una distribución binomial. Se ha utilizado este tipo de modelo teniendo en cuenta que los datos no puede ser considerada normal. Los dos factores se comparan de forma simultánea y se utiliza el ID de usuario como factor agrupante. Se ha realizado un análisis para el caso de la detección del estímulo (si el

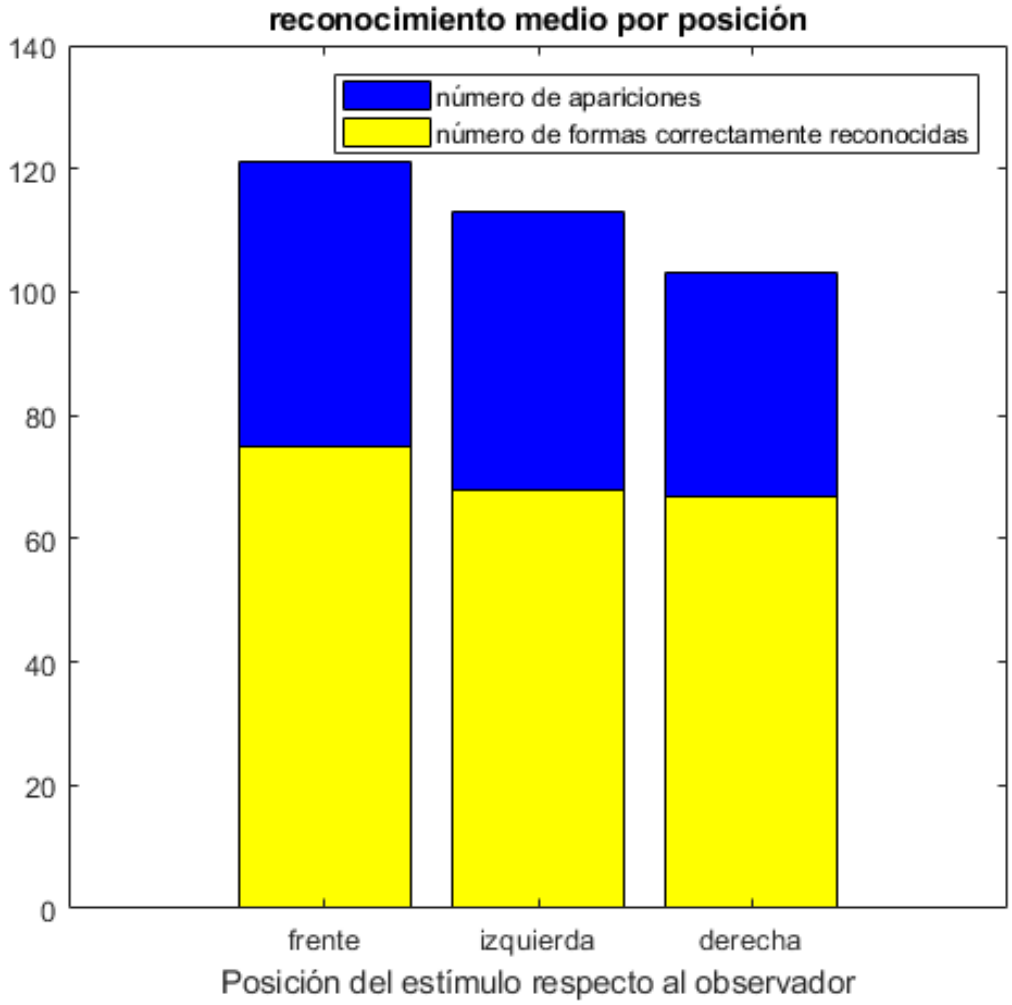


Figura A.3: En azul detección sobre el total de estímulos y en amarillo reconocimiento para cada posición del estímulo visual.

participante ha visto algo o no) y otro para el reconocimiento del estímulo (si ha sido capaz de distinguir correctamente la forma que ha visto o no).

En el caso de la detección de estímulos, ni la forma (p-valor=0.63) ni la posición (p-valor=0.21) del estímulo ni las interacciones de primer nivel entre ambos (p-valor=0.27) influyen significativamente en la detección del mismo. Por el contrario, en el caso del reconocimiento de estímulos la forma sí que influye significativamente (p-valor=0.001), mientras que la posición (p-valor=0.28) del estímulo y la interacción de ambos factores (p-valor=0.11) siguen sin hacerlo.

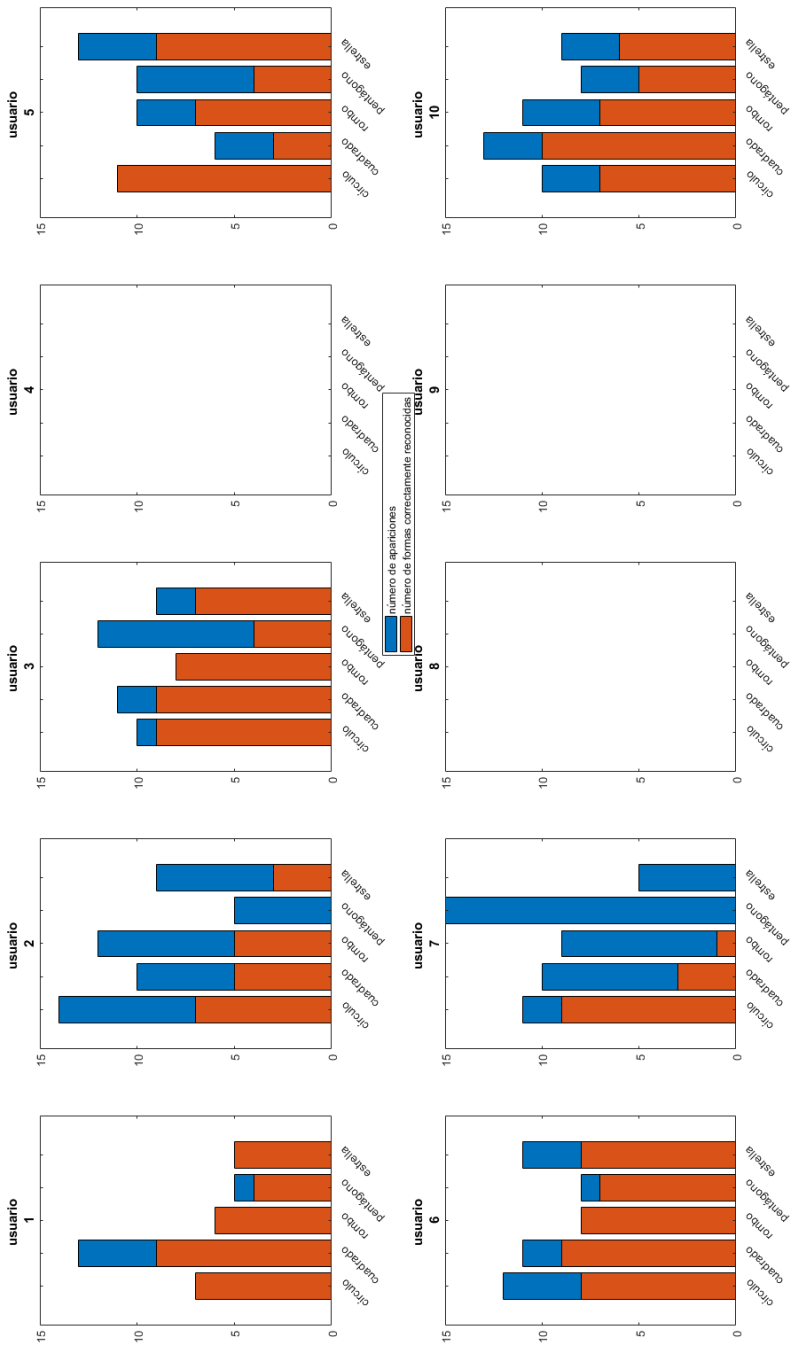


Figura A.4: En azul detección sobre el total de estímulos y en amarillo reconocimiento para cada forma del estímulo visual.

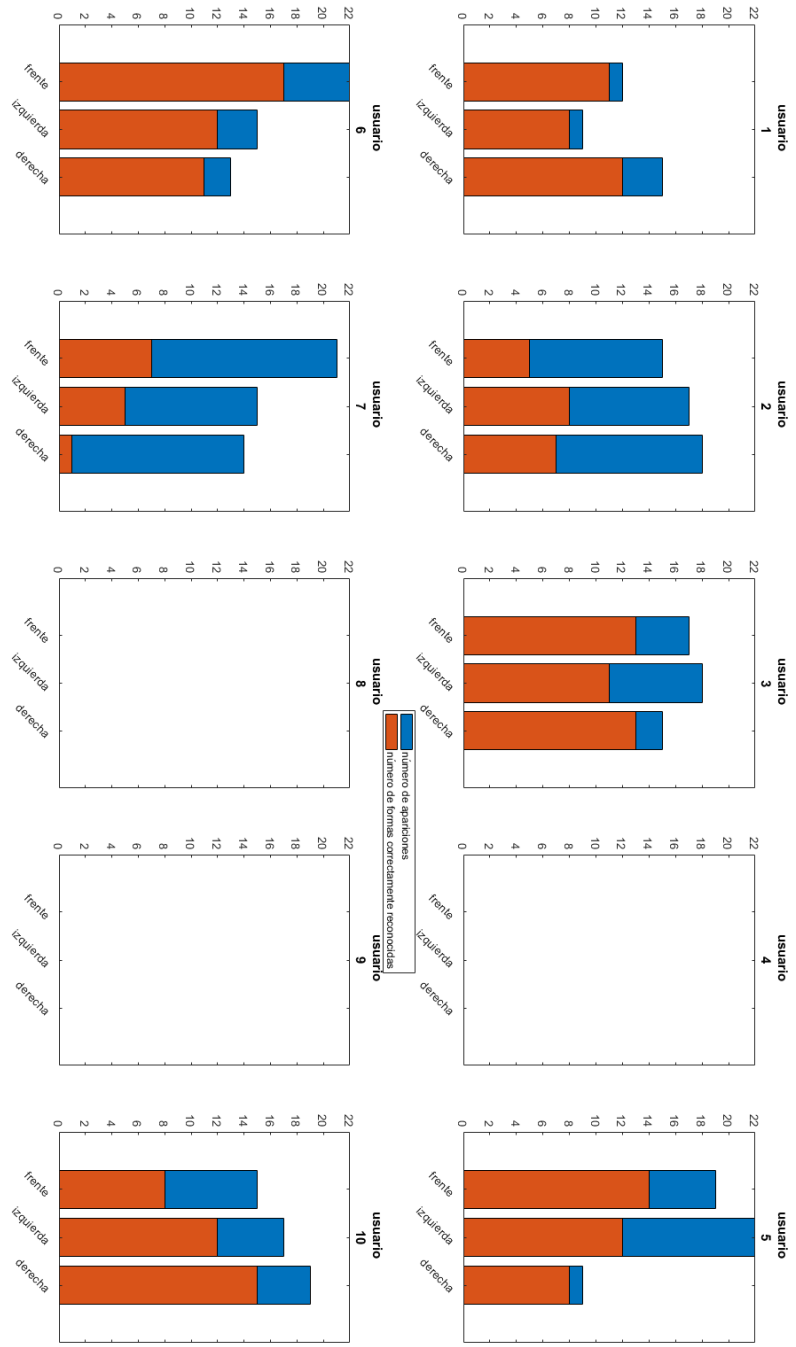


Figura A.5: En azul detección sobre el total de estímulos y en amarillo reconocimiento para cada posición del estímulo visual.

A.3. Conclusiones

Como se esperaba, se ha obtenido un nivel de detección alto para los estímulos visuales en las tres localizaciones utilizadas. El reconocimiento de la forma visual, con un porcentaje menor de aciertos, provee de un nivel de dificultad adicional a la tarea presentada.

Se puede concluir que los estímulos son visibles en las condiciones y escenario presentados en este experimento. Hay que tener en cuenta que el usuario se mueve libremente por la escena. Sus movimientos (tanto cambios de posición como de orientación) influyen en cierto grado en los estímulos que no son detectados. Es posible que algunos estímulos hayan aparecido cuando el usuario estaba girando la cabeza, o incluso parpadeando. Por ello, es lógico que la tasa de detección no siempre sea un 100 %.

Teniendo esto en cuenta, no se debe esperar una tasa de aciertos del 100 % en la condición visual del experimento descrito en la Sección 4.1, pero sí una diferencia significativa entre las tasas de acierto de la condición visual y la audiovisual.

Anexo B. Formulario participantes

En este anexo se presenta el formulario que han rellenado todos los participantes de los dos experimentos llevados a cabo en este trabajo. La primera cara (pre-test) se rellenaba antes de comenzar el experimento. La segunda cara (post-test) se realizaba al finalizar el experimento. Se animaba a los participantes a escribir cualquier apreciación que considerasen interesante sobre el experimento en esta parte del cuestionario.

ID participante: _____ Fecha: __/__/__

Pre-test

Datos sobre el participante

Edad: _____ Género: _____

Trabajo:

Estudios:

¿Padece algún problema de visión?: Sí No
En caso afirmativo, cual:
En caso afirmativo, ¿utiliza algún tipo de corrección para el mismo?
(Gafas, lentillas, etc.): Sí No

¿Ha utilizado alguna vez unas gafas de realidad virtual (VR)?: Sí No
En caso afirmativo, ¿con qué frecuencia?:
Una vez Ocasionalmente Regularmente
En caso afirmativo, ¿alguna vez ha padecido malestar al usar las gafas de VR?:
Sí No Síntomas:

En caso afirmativo, ¿Ha experimentado algún otro tipo de problema al utilizar las gafas de VR?: (Desfase entre audio y vídeo, cortes en la imagen, etc.):
Sí No Cuales:

Post-test

Datos sobre el experimento

¿Siente algún tipo de malestar?: Sí No

En caso afirmativo, ¿cuál?:

(Mareo, fatiga, dolor de cabeza, dolor de ojos, incomodidad, nauseas, etc.)

Durante el experimento, ¿ha notado algún tipo de cambio a lo largo del tiempo?:

Durante el experimento, ¿ha visto u oído algo extraño o fuera de lugar?:

¿En algún momento ha sido consciente del tema o los temas que se trataban en la radio que se oía de fondo? En caso afirmativo, indique a grandes rasgos lo que recuerde:

¿Hay algo más que quiera comentar sobre el experimento?: