



**Universidad**  
Zaragoza

## Trabajo Fin de Grado

**Edición en Realidad Virtual:  
influencia de técnicas sonoras en la continuidad narrativa**

**Edition in Virtual Reality:  
influence of sound techniques on narrative continuity**

Autor:  
**Javier Camón Julián**

Directora: **Ana Serrano Pacheu**

Ponente: **Diego Gutiérrez Pérez**

Grado en Ingeniería de Diseño Industrial y Desarrollo de Producto (EINA)  
Septiembre de 2017

 **Graphics and  
Imaging Lab**



**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).

### TRABAJOS DE FIN DE GRADO / FIN DE MÁSTER

D./Dª. Javier Camón Julián,

con nº de DNI 77132645Y 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) Grado (Título del Trabajo)

Edición en realidad virtual: influencia de técnicas sonoras en la continuidad narrativa

---

---

---

---

---

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

Zaragoza, 18 de Septiembre de 2017

Fdo: Javier Camón Julián



# Resumen

---

La realidad virtual (RV) es un reciente avance tecnológico que está abriendo las puertas a nuevas experiencias audiovisuales y procesos de diseño mucho más inmersivos que los convencionales. Su aplicabilidad al mundo del cine, los videojuegos y los entornos virtuales hace gala de un enorme potencial, dada su capacidad de situar al usuario, ya sea espectador pasivo o agente activo, en un mundo virtual diseñado a voluntad del creador.

Este proyecto se enfoca en la adaptación de la cinematografía tradicional a la realidad virtual. Más concretamente se centra en analizar la influencia del sonido panorámico direccional (o *panning*) en la percepción de la continuidad cinematográfica en este nuevo entorno, con el fin de obtener una serie de conclusiones que permitan diseñar un marco sonoro que facilite la interacción audiovisual y la transmisión de la intención del director o editor.

Con esta meta se desarrollan y realizan una serie de experimentos de tipo prueba de usuario, en los que se somete a los sujetos a escenas en realidad virtual complementadas con sonido panoramizado y orientable, después de lo cual se comparan los resultados con los obtenidos tras los mismos experimentos realizados sin la panoramización del sonido.

Se elaboran entonces una serie de conclusiones a partir del conocimiento adquirido, cuya motivación será mostrar qué funciona mejor para contar una historia en realidad virtual: es decir, establecer ciertas pautas del lenguaje cinematográfico en RV, arrojando así algo de luz en la cada vez más débil reinante oscuridad en el mundo del cine en este nuevo entorno.



# Abstract

---

Virtual reality (VR) is a recent computer technology that is unfolding the possibilities of new audiovisual experiences and more immersive design processes than we are used to. Its applicability to the cinematography world, videogames and virtual environments flaunts a huge potential, given its power to ship the user to a virtual world designed by someone else.

This project is focused on the adaptation of the traditional cinematography to VR. More specifically, it pays special attention to the analysis of the influence of panoramic-orientated sound (*panning*) on the perception of cinematographic continuity in this new environment, with the purpose of achieving some conclusions that allow us to design a sound framework that softens the audiovisual interaction and eases the communication of the creator's intentions and ideas.

With this goal in mind, a series of user-test experiments are developed and performed, in which users are subjected to panned and orientable sound VR scenes, after which the results are compared to those obtained after same experiments performed without panning the sound.

Then, a set of conclusions are drawn from the acquired knowledge, whose motivation will be to show what works better to tell a story in virtual reality: that is, to establish some patterns of the VR language, thus slightly throwing a bit of light on the prevailing but decrescent darkness in the world of VR cinema.



# Agradecimientos

---

Me gustaría dar las gracias...

A Ana Serrano, directora del proyecto, por el trabajo, la dedicación y la paciencia que ha tenido conmigo a lo largo de tantos meses.

A Belén Masiá, por sus enseñanzas, implicación y su predisposición a ayudar siempre que fuera posible.

A Diego Gutiérrez, por su compromiso, implicación e interés, y por darme la oportunidad de trabajar en este proyecto y con este grupo tan impresionante.

A Jaime Ruiz-Borau, por su legado y su disposición a echarme una mano cuando fuera necesario; a Sandra, Víctor e Ibón, por su ayuda con el código, y a todos los compañeros del Graphics and Imaging Lab por sus consejos y por contribuir a crear un ambiente muy bueno.

A los amigos y todas aquellas personas que han dejado lo que estaban haciendo en la otra punta de Zaragoza (o de España) para venir a participar en los experimentos.

A mi familia, en especial mis padres y mi hermana, que jamás se cansarán de apoyarme; y a Laura, que ha empujado como nadie.



# Índice

---

Resumen .....	5
Abstract .....	7
Agradecimientos .....	9
Índice .....	11
Índice de tablas y figuras .....	13
1. Introducción.....	15
1.1 Contexto general .....	15
1.2 Objetivos .....	18
1.3 Planificación .....	18
2. Estudio previo y motivación .....	21
2.1 Cinematografía tradicional .....	21
2.2 Cinematografía en realidad virtual.....	23
2.3 Sonido panoramizado con orientación, o <i>panning</i> .....	26
2.4 Edición cinematográfica: J-cut .....	28
2.5 Motivación e hipótesis.....	30
3. Experimentos .....	31
3.1 Conceptos involucrados.....	31
3.2 Primer experimento: sonido con <i>panning</i> orientativo.....	33
3.2.1 Planteamiento.....	33
3.2.2 Preparación.....	34
3.2.3 Realización del experimento .....	39
3.2 Segundo experimento: J-cut .....	40
4. Entornos de trabajo .....	41
4.1 Entorno audiovisual .....	41
4.1.1 Oculus Development Kit 2.....	41
4.1.2 Unity3D .....	42
4.1.3 Adobe Premiere CC2016.....	44

4.2 Entorno de adquisición de datos.....	44
4.2.1 PupilCapture 0.8.7 .....	44
5. Análisis de datos .....	47
5.1 Estudio de las métricas.....	47
5.2 Análisis.....	48
6. Resultados.....	49
6.1 Primer experimento .....	49
6.1.1 Fotogramas en alcanzar una <i>ROI</i> .....	49
6.1.2 Error en la trayectoria en la búsqueda de ROIs .....	50
6.2 Segundo experimento.....	51
6.2.1 Fotogramas en alcanzar una ROI .....	51
6.2.2 Error en la trayectoria en la búsqueda de ROIs .....	51
7. Conclusiones.....	53
7.1 Implicaciones para el futuro .....	54
Bibliografía.....	56
Anexos .....	59
A. Recopilación de vídeos 360 o RV: enlaces .....	60
B. Clips de video 360 representativos .....	62
C. Procesado de datos.....	63
C.1 Recopilación de datos .....	63
C.2 Pre-procesado: baseline.....	63
C.3 Tratamiento de los datos.....	65
D. Encuesta realizada en el experimento .....	72
E. Espacio de variables .....	73
F. Código desarrollado .....	74

# Índice de tablas y figuras

---

1.1 Simulación de lo que ve un usuario de realidad virtual	15
1.2 Componentes de un dispositivo de realidad virtual	16
1.3 Planificación del trabajo	19
1.4 Diagrama de Gantt	20
2.1 Escena de <i>Paprika</i> , de Satoshi Kon	23
2.2 El zootropo y el praxinoscopio	23
2.3 Estructura de secuencias donde los colores son las escenas de 360° y las ROIs son los puntos	25
2.4 Ejemplo de mapa de saliencia	26
2.5 Funcionamiento del panning	27
2.6 Localización de la fuente de un sonido mediante triangulación	28
2.7 Esquema de un J-cut durante la edición	28
2.8 Ejemplo de J-cut	29
3.1 Diagrama del ojo humano	33
3.2 Las cuatro localizaciones de las escenas	33
3.3 Esquema de los dos desalineamientos	35
3.4 Configuraciones de las ROIs	36
3.5 Combinaciones de las ROIs con los desalineamientos	36
3.6 Distribución de los altavoces virtuales siguiendo a los personajes	37
3.7 Detalle del eyetracker	38
3.8 Montaje del sistema para el experimento	39
4.1 Detalles del casco HMD de Oculus	42
4.2 Interfaz de Unity3D en Windows 8	42
4.3 Objeto esfera donde se proyecta la imagen 360°	43
4.4 Interfaz de Adobe Premiere CC 2017	44
4.5 Detalle del montaje del <i>eyetracker</i> en el Oculus DK2	45
4.6 Ojo del sujeto en el interfaz del programa PupilCapture en Ubuntu	46

6.1 Comparativa de los resultados de Fotogramas en alcanzar una ROI (A)	49
6.2 Comparativa de los resultados de Fotogramas en alcanzar una ROI (R)	50
6.3 Comparativa de los resultados de Error en la trayectoria en la búsqueda de ROIs frente al desalineamiento y el uso de sonido con panning orientativo	50
6.5 Comparativa de los resultados de Fotogramas en alcanzar una ROI frente a su desalineamiento y a la baseline sin J-cut.	51
6.6 Comparativa de los resultados de Error en la trayectoria en la búsqueda de ROIs frente al desalineamiento y el uso de sonido con panning orientativo	51

# 1. Introducción

---

## 1.1 Contexto general

Aparentemente, los términos *real* y *virtual* representan ideas opuestas y contradictorias, ya que si algo es real no puede ser virtual, y viceversa. La conjunción de ambos, sin embargo, adquiere un nuevo sentido: la *realidad virtual* consiste en un entorno audiovisual digital que pretende sustituir al real haciendo percibir al usuario que se encuentra *inmerso* en un mundo distinto al que está físicamente, como se puede ver en la Figura 1.1; ya sea creado a partir de escenas grabadas en escenarios reales o virtuales (cinematografía), interactivo, o generado artificialmente por un ordenador (videojuegos, simuladores y programas de diseño).

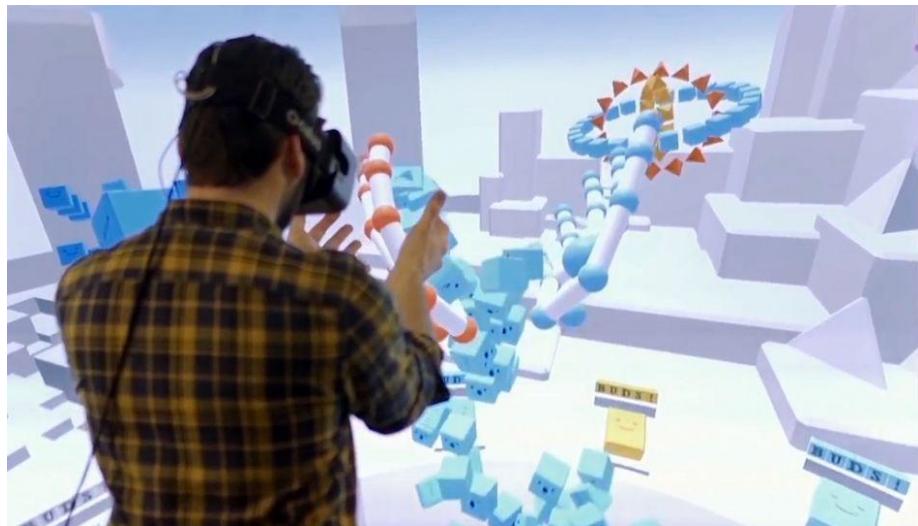


Figura 1.1 Simulación de lo que ve un usuario de realidad virtual (Fuente: fundaciontelefonica.com [1])

La *realidad virtual (RV)* ha sufrido una evolución notable a lo largo de los últimos años, haciendo viable su aplicación a los videojuegos y al cine. La mejora en los sensores, la resolución de las pantallas y el incremento de potencia de los nuevos ordenadores han sido claves en el desarrollo de esta tecnología, si bien continúa en estado de refinamiento

## 1. Introducción

---

constante gracias a diversas marcas de desarrollo tanto de software como de hardware, como son Oculus o HTC, que han ido lanzado sendos productos cada vez más completos y competitivos -Oculus Rift y Vive, respectivamente, que superan los 400€. Por otro lado, marcas como Samsung están desarrollando carcasa para utilizar el smartphone como pantalla, como el llamado Gear VR, rebajando así el precio a unos 130€.

Según la compañía de información IDC, los ingresos mundiales en el mercado de la realidad virtual superarán en 2017 los 14.000 millones de dólares, pudiendo acercarse a los 150.000 millones para 2020. Las empresas que están invirtiendo en ello son muy diversas: Facebook, HTC, Rothenberg Ventures, Disney, Samsung o Intel (Fuente: Innovation Enterprise [2]).

Los dispositivos de realidad virtual cuentan principalmente con un casco *HMD* (*Head Mounted Display*) conectado al ordenador, pudiendo además incorporar auriculares, mandos de control y cámara de seguimiento. El casco *HMD* es el responsable de mostrarnos la imagen correspondiente en función de la orientación de nuestra cabeza, gracias a los sensores de orientación (y también de posición para entornos simulados) incorporados en este. La pantalla muestra a cada ojo la imagen necesaria para que nuestro cerebro monte la escena: podemos ver a nuestro alrededor como si nos encontráramos en el emplazamiento mismo, eligiendo la dirección en la que queremos mirar en todo momento. Los auriculares emiten el sonido correspondiente a la situación en la que nos encontramos, y los mandos nos permiten controlar otros parámetros, como los gestos y otros movimientos hechos con las manos, o acciones adicionales. La Figura 1.2 muestra estos componentes durante su uso.



Figura 1.2 Componentes de un dispositivo de realidad virtual (Fuente: Oculus [3])

La enorme utilidad de la realidad virtual en el mundo del diseño queda patente no solamente en el nicho de los videojuegos, en el que efectivamente está demostrando tener un

increíble potencial, sino también como herramienta de diseño asistido por ordenador -DAO, o *CAE*, *Computer Aided Design*, en inglés-, y de fabricación asistida por ordenador -CAM, *Computer Aided Manufacturing*-. Una encuesta de PwC muestra que el 38.8% de las compañías están utilizando actualmente la RV o la RA (realidad aumentada) para el diseño y desarrollo de producto, siendo el entrenamiento de seguridad y de ensamblaje virtual los ámbitos más explotados. Se suele enfocar igualmente a la mejora y refinamiento de los procesos de diseño, como la monitorización de la calidad del acabado (Fuente: Innovation Enterprise [2]).

Pero el desarrollo de la realidad virtual también ha traído consigo la posibilidad de adaptar el cine tradicional a estos dispositivos. Durante más de un siglo, la cinematografía ha ido explorando la forma de transmitir realismo y naturalidad, y como consecuencia se han ido elaborando una serie de reglas que buscan un flujo continuo en la cognición del espectador y funcionan como guía a la hora de editar un montaje. Esta *edición en continuidad* (Bordwell et al. 1997 [4]) asegura que los distintos planos y escenas que corresponden a una sola acción o evento se perciban como tales, y existe una coherencia tanto espacial como temporal en el flujo de imágenes visualizadas: es decir, que se pueda entender como una historia.

Y si es precisamente de la necesidad y el deseo de contar una historia, de compartir una experiencia, una vivencia, o una ocurrencia, de donde nace la cinematografía, igualmente ocurre con el uso de la realidad como plataforma para comunicarnos. Aquel que busque una forma nueva, más inmersiva, de transmitir un contenido, podrá encontrar en la realidad virtual la herramienta perfecta. Numerosos son los artistas y los estudios creativos, como por ejemplo *Felix & Paul*, que se están lanzando hoy en día a desarrollar contenido para RV: el mundo de la experiencia audiovisual empieza a abrirse a esta tecnología en un formato del que no se sabe mucho todavía.

Tal es la experiencia que ofrece este medio que su popularidad comienza a alcanzar cotas impensables hace una década; sin ir más lejos, por primera vez, este año 2017, la Academia de Artes y Ciencias Cinematográficas de Hollywood ha nominado a *Pearl*, un cortometraje en formato “video 360” -compatible para su visualización en RV- dirigido por Patrick Osborne para *Google Spotlight Stories*, a los Premios Oscar. Este mismo año un conocido certamen de cine fundado por Robert de Niro, el *Tribeca Film Festival*, acepta la participación de productores en RV en el acto principal del evento, e incluso podrán ser galardonados. Por otra parte, como muestra de un creciente interés por esta plataforma el 70º Festival de Cannes proyectó en RV el corto *Carne y arena*, del director Alejandro G. Iñárritu; de igual modo, el 50º Festival de Sitges contará con contenidos 360º de corte fantástico, y La Mostra de Venecia será uno de los primeros certámenes de clase A en organizar una competición entre 22 filmes en realidad virtual.

Por otro lado, florecen nuevos festivales con la cinematografía para realidad virtual (y aumentada) como principal protagonista, como el canadiense FIVARS (*Festival of International Virtual and Augmented Reality Stories*), el español *Movistar Barcelona 360 VR Fest*, o el mexicano *VR Fest MX*.

## 1. Introducción

---

### 1.2 Objetivos

Actualmente, la creación de contenido en entornos virtuales goza de un gran auge, pero existe una carencia de pautas sólidas que muestren cómo se debe hacer la adaptación del cine tradicional a la realidad virtual, debida sin duda a la falta de experiencia: qué técnicas y recursos funcionan, cuáles no lo hacen, y qué nuevas posibilidades ofrece este nuevo medio. Dado que el formato audiovisual es distinto, puesto que se graba en vídeo un entorno de 360° a nuestro alrededor y el observador tiene la potestad para mirar en la dirección que quiera, de nada sirve ahora controlar la orientación de la cámara o elegir los planos que han de ser vistos a la hora de editar el montaje; será necesario establecer nuevas reglas de edición a partir de resultados empíricos obtenidos con observadores reales sometidos a pruebas de visualización de contenido en RV.

En este trabajo nos enfocamos, en particular, en conocer cuál debe ser el uso del sonido cuando se edita contenido en realidad virtual; nuestro objetivo es indagar qué ocurre con la continuidad narrativa percibida en los cortes cinematográficos y cómo el sonido panoramizado (o con *panning*) afecta a la atención del usuario, dada la ausencia de información acerca del efecto de los diferentes recursos sonoros en la transmisión de continuidad narrativa en este contexto. Para realizar la investigación emplearemos un método heurístico: se buscará obtener una serie de conclusiones sobre el problema abordado y diseñando para ello unos experimentos basados en pruebas de usuario. Se trata de un trabajo de investigación, por lo que se llevará a cabo un análisis teórico inicial, se generarán hipótesis acerca del problema, se realizarán los tests para verificar dichas hipótesis y finalmente se analizarán los resultados, lo que nos conducirá a elaborar las conclusiones.

### 1.3 Planificación

Durante la realización de este trabajo se suceden las siguientes etapas:

- En primer lugar se aborda el enfoque del problema, la planificación y la proyección del proceso de resolución.
- Se lleva a cabo un estudio de las técnicas y recursos cinematográficos tradicionales, en especial aquellas que estuvieran relacionadas con el sonido y la percepción de la continuidad narrativa en los cortes entre escenas.
- Se realiza también un estudio del estado del arte en realidad virtual, haciendo hincapié en la cinematografía adaptada a la realidad virtual.
- Se lleva a cabo un análisis teórico de los factores relacionados con el sonido relevantes en la percepción de continuidad cinematográfica en realidad virtual.

- A raíz del análisis se diseñan dos experimentos de prueba de usuario con el fin de evaluar dichos factores y obtener información acerca de su influencia en la continuidad cinematográfica en RV. Para crear los estímulos se utilizan vídeos en formato 360º grabados por una empresa local y se editan con Premiere.
- Se llevan a cabo los experimentos con un total de 60 sujetos en total para los dos, 35 el primero y 25 el segundo. Para realizar los experimentos se utilizarán el software Unity3D y el dispositivo de realidad virtual OculusDK2, además del software PupilCapture.
- Se procesan los resultados con Matlab y se analizan con SPSS, programas de software de tratamiento de datos y estudio estadístico, respectivamente.
- Se generan las conclusiones. Se reflexionan ciertas implicaciones y consideraciones para el futuro.

En un primer momento se plantea el trabajo para 4 meses de duración, comenzando en marzo de 2017 (ver Figura 1.3). Sin embargo, debido a un proceso de desarrollo y realización de los experimentos más lento de lo esperado, debido a los obstáculos encontrados y a la dependencia de la asistencia de sujetos a las pruebas, se prolonga hasta mediados de septiembre (ver Figura 1.4).



Figura 1.3 Planificación del trabajo

## 1. Introducción

---



Figura 1.4 Diagrama de Gantt

## 2. Estudio previo y motivación

---

### 2.1 Cinematografía tradicional

El cine tradicional se crea, en la mayoría de los casos, editando y empalmando las tomas en un flujo discontinuo de planos separados por los *cortes*; de lo contrario, una película duraría lo mismo que el tiempo que transcurre en ella, y no se podrían hacer cambios instantáneos en la localización de la acción. Existen algunos ejemplos de excepciones a esta regla, como *La Soga* (Alfred Hitchcock, 1948 [5]) y *Birdman* (Alejandro G. Iñárritu, 2014 [6]), en las que se graban *planos-secuencia* (planos sin cortes en los que la cámara se mueve siguiendo a la acción) magistralmente empalmados de forma que aparentan ser uno sólo, o *Hijos de los Hombres* (Alfonso Cuarón, 2006 [7]), en la que se suceden largos planos-secuencia en momentos determinados de la película para transmitir ciertas sensaciones. Sin embargo, lo más habitual es hacer uso de una gran variedad de tipos de cortes no sólo por necesidad, sino también por el deseo de causar efectos concretos en el espectador.

Por otro lado, entendemos *continuidad* como la unión de las partes que forman un todo que se desarrolla en el tiempo, de forma cohesionada y coherente. Esta definición parece contrastar con la edición cinematográfica basada en los cortes, pero sin embargo, pese a esta fragmentación de la línea espacio-temporal, los espectadores somos capaces de unir de forma automática las escenas de forma coherente según lo que se denomina *continuidad narrativa*. Para comprender cómo los seres humanos percibimos y entendemos dicha continuidad de los eventos a pesar de la edición se han realizado diversos estudios cognitivos basados en la teoría de segmentación de eventos.

Según Smith et al. [8] una posible explicación de por qué la edición en continuidad funciona es que “produce una corriente de información audiovisual que se parece lo suficiente a nuestra experiencia del mundo real como para percibirla como tal, sin requerir habilidades cognitivas especiales” (Anderson [9]; Bordwell et al. [10]; Cutting [11]; Gibson y Rader [12]). De hecho, aunque algunas técnicas cinematográficas sean mera convención, como el difuminado de un flashback, la mayoría derivan de acciones existentes en nuestra vida cotidiana, como el fundido en negro.

## 2. Estudio previo y motivación

---

Smith afirma que la continuidad es considerada por muchos como “una ilusión creada en la mente del observador y no una cualidad inherente al estímulo (o sea, la película)”, por lo que se ha de centrar el estudio en la relación entre el contenido y el espectador. Experimentos realizados con cortes de prueba presentados a sujetos (Levin & Simons [13]) determinaron que la atención juega un rol esencial en la percepción de la continuidad, demostrando que sólo concentrándonos en buscar algo concreto que sabemos de antemano que está ahí somos capaces de percibirlo: es lo que se llama *atención selectiva*. El famoso experimento del gorila de Chabris y Simons de 1999 [14] lo ejemplifica: se les pide a los sujetos que cuenten el número de veces que se pasan la pelota los personajes de un vídeo, y en un determinado momento un gorila pasa por detrás de ellos, pero sólo un pequeño porcentaje de los sujetos lo ve. Por otro lado, en una situación sin riesgo tendemos a relajar nuestra atención y a dejar que nuestro cerebro se encargue de unir los cables sueltos de forma automática, siendo este un mecanismo de supervivencia desarrollado a lo largo de nuestra evolución.

Varios estudios empíricos (Zacks y Swallow [15], Zacks et al. [16]) avalan la teoría de la segmentación de eventos, concordando en que este proceso es un componente automático de nuestro procesamiento de la información percibida. Nuestro cerebro asimila mejor esta información si está fragmentada en eventos discretos, y gracias a esto será capaz de predecir más fácilmente el futuro próximo. Cuando algo no sucede como era de esperar, se creará un nuevo evento, de forma que los cambios inesperados marcarán el punto de inflexión en el flujo de acontecimientos. Una de las conclusiones de Magliano y Zacks [17] más relevantes es que cuando los cortes conllevan un cambio notable, el cerebro se adapta al cambio construyendo una nueva imagen de la situación, en lugar de intentar explicar la brecha en la continuidad narrativa (Zacks et al. [16]; Smith [18]; Zacks et al. [19]). Teniendo esto en cuenta se puede entender mejor por qué los cortes, aunque coinciden con las fronteras de los eventos de una historia, no perjudican a la continuidad narrativa.

No obstante, los cortes no se bastan para conformar todo el lenguaje cinematográfico. A lo largo de la historia del cine se han ido desarrollando numerosos recursos que permiten al director y al editor transmitir su intención a través del montaje, además de las ya mencionadas pautas de edición. Por ejemplo, utilizando distintos tipos de planos se puede controlar la evolución de la atención del espectador según Smith [18]: si se la desea dirigir hacia un personaje concreto, después de un *plano general* se muestra un *plano medio* y luego un *plano sobre el hombro*. Se puede enfatizar una discusión mediante el *plano-contraplano*, presentar o despedir personajes con el *plano de salida/entrada*, o seguir los pasos de alguien con un *plano de barrido* [20, 21]. También hay movimientos que la cámara no debe seguir si se quiere mantener intacto el sentido de orientación del espectador, como la regla de los 180° que se ejemplifica en Paprika (Satoshi Kon, 2006 [22]), como muestra la Figura 2.1. Estas pautas y recursos se han ido descubriendo mediante la experimentación, a través de ensayo y error, a veces incluso transmitiendo su éxito o fracaso con la propia película. Son convenciones que pueden quedar refutadas o aceptadas con la respuesta del público, y el tiempo dicta la relevancia de cada una.



Figura 2.1 Escena de *Paprika*, de Satoshi Kon, que explica la regla de los 180º [22].

## 2.2 Cinematografía en realidad virtual

Con la invención de la cinematografía en la última década del siglo XIX lo que se pretende es adaptar un guión o una historia -ya estuviera escrita o en imágenes- a un formato visual que se asemeje más a nuestra percepción de la realidad como un continuo temporal, y poder transmitir el movimiento con mayor naturalidad -con esa idea se habían creado el estroboscopio, el zoóptropo, el praxinoscopio y el cronofotógrafo entre otros (ver Figura 2.2), y finalmente el cinematógrafo. Contar una historia en realidad virtual supone un salto de orden similar al acaecido con la creación del cine, pues así como con estas herramientas se consiguió transmitir el contenido visual de forma más realista, con la realidad virtual se logra otro hito más allá, al no sólo transmitir imágenes en movimiento, sino conseguir que esas imágenes sustituyan por completo la total percepción de nuestro entorno a tiempo real.

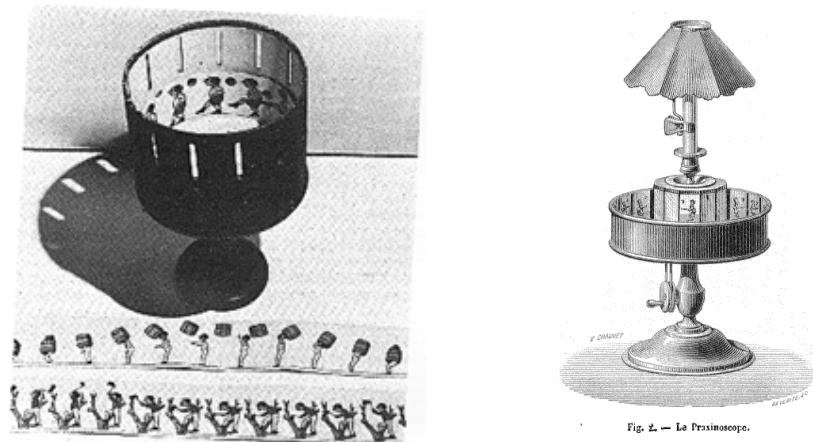


Figura 2.2 El zootropo y el praxinoscopio (Fuente: blog de E. Martínez-Salanova [23])

## 2. Estudio previo y motivación

---

Y así como el cine, según Sebastian Sylwan -socio creativo y director tecnológico de la mayor productora de realidad virtual del mundo, Felix & Paul- [24], sufrió desde sus inicios un desarrollo creativo a la par que tecnológico, lo mismo ha de ocurrir en la actualidad:

*“A principios del siglo XX los estudios tenían que inventar sus propias cámaras, sus técnicas de procesado del celuloide, de iluminación de las escenas. La innovación tecnológica era total. Incluso tenían que pensar cómo presentar las obras al público, cómo distribuirlas. Todo eso es algo que, aun estando en un mundo completamente digital, hay que plantearse en la realidad virtual.”*

Cuando miramos una escena de una película en el cine o en la televisión, se nos muestra toda la información disponible únicamente dentro del encuadre de la cámara, que ha sido diseñado para satisfacer los requisitos de la historia. Por tanto, la libertad de observación del usuario se ve totalmente restringida de antemano, de la misma forma que un escritor describe únicamente lo que le interesa para recrear el escenario en la mente del lector. De hecho, como ya se ha explicado, existen ciertas técnicas para buscar un efecto u otro en el observador: distintos tipos de planos, movimientos de cámara, enfoques, cortes y efectos sonoros. En cambio, en realidad virtual la situación cambia drásticamente, al ser el usuario libre de elegir la región de la escena que va a mirar. Esto supone un gran problema a la hora de narrar una historia, y surge la pregunta de cómo recuperar ese control sobre lo que ve el espectador, que se pierde al darle la libertad de elegir la orientación. Porque no interesa que el usuario se pierda ningún hecho o detalle relevante, pero ¿podemos dirigir su atención?

Así lo explica Sebastian Sylwan [24]:

*“No podemos simplemente llegar y adaptar lo que sabemos del cine a la realidad virtual. El corte, por ejemplo, tiene un efecto distinto. El cine está muy establecido en la narrativa, vas de un plano a otro sin problema. Pero aquí tiene una semántica distinta: eres tú el que estás en un lugar y, sin ningún preaviso ni ninguna acción por tu parte, inmediatamente cambias de sitio.”*

Numerosas ideas y pruebas se han realizado con el fin de encontrar la mejor forma de llevar de la mano al observador a lo largo de la narración: Jessica Brillhart, cineasta y antigua responsable de Google VR, [25] propone, introduciendo el concepto de *regiones de interés* (*ROI* en inglés) como las zonas de la imagen en 360º donde transcurre la acción más relevante para la historia, una nueva estructura de secuencias en la que se alinean dichas ROIs consecutivamente de forma espacio-temporal, presuponiendo dónde va a mirar el espectador en cada momento, como muestra la Figura 2.3. Funciona de tal forma que a un observador que esté mirando una ROI le sería mostrada tras el corte, en la misma dirección, la ROI de la siguiente escena, evitando así la pérdida de la atención, y por tanto, de la transmisión de la información.

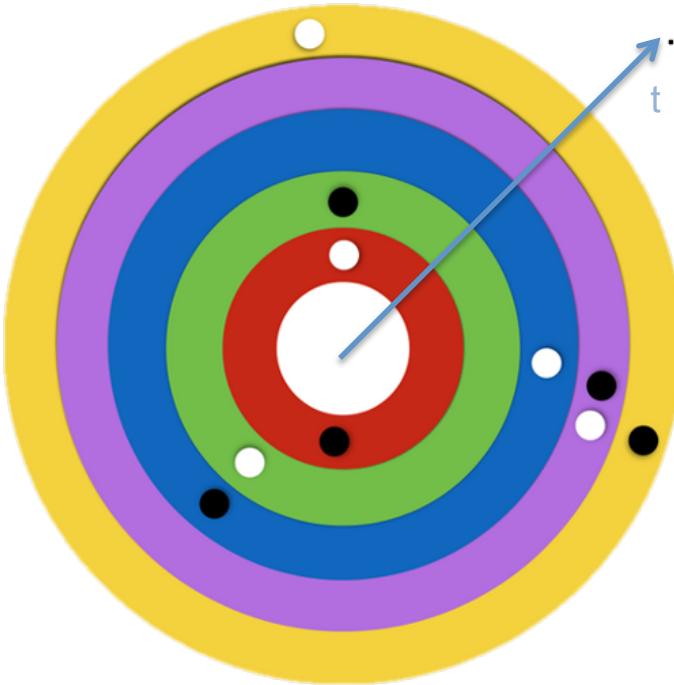


Figura 2.3 Estructura de secuencias donde los colores son las escenas de 360º y las ROIs son los puntos. Se debe entender la línea temporal con inicio en el centro hacia el exterior (Fuente: blog de Jessica Brillhart [25])

Para conocer estas zonas será necesario investigar el comportamiento y la atención del público mediante *mapas de saliencia*, representaciones típicamente a color sobre la imagen que muestran la capacidad de llamar la atención de cada punto de la misma. Se puede observar un ejemplo en la Figura 2.4; en este caso, como es habitual, se ha utilizado una escala cromática que le confiere la apariencia de un *mapa de calor*: los lugares más rojizos son los más observados, mientras que los verdes son los que menos han llamado la atención del observador. Esto se consigue medir mediante seguidores oculares o *eyetrackers*, que registran la posición de las pupilas y calculan la zona a la que mira el sujeto. Los mapas de saliencia son una forma de estudiar la segmentación de imágenes, utilizados ampliamente en el mundo de la mercadotecnia y de la publicidad.

Brillhart [25] afirma que para hacer cine en RV es necesario deshacerse de la idea de marco cinematográfico, pues hará plantearse las preguntas equivocadas a los cineastas. No se puede forzar a un espectador a mirar a una zona concreta, pero sí que se puede estudiar su comportamiento para predecir dónde va a mirar, qué hacer para llamar su atención, o animarle a mirar en una determinada dirección. Sabemos, por ejemplo, que cuando entramos en una estancia nueva para nosotros y algo importante está ocurriendo en el centro, no nos ponemos a inspeccionar los espacios recónditos ni las esquinas.

## 2. Estudio previo y motivación

---

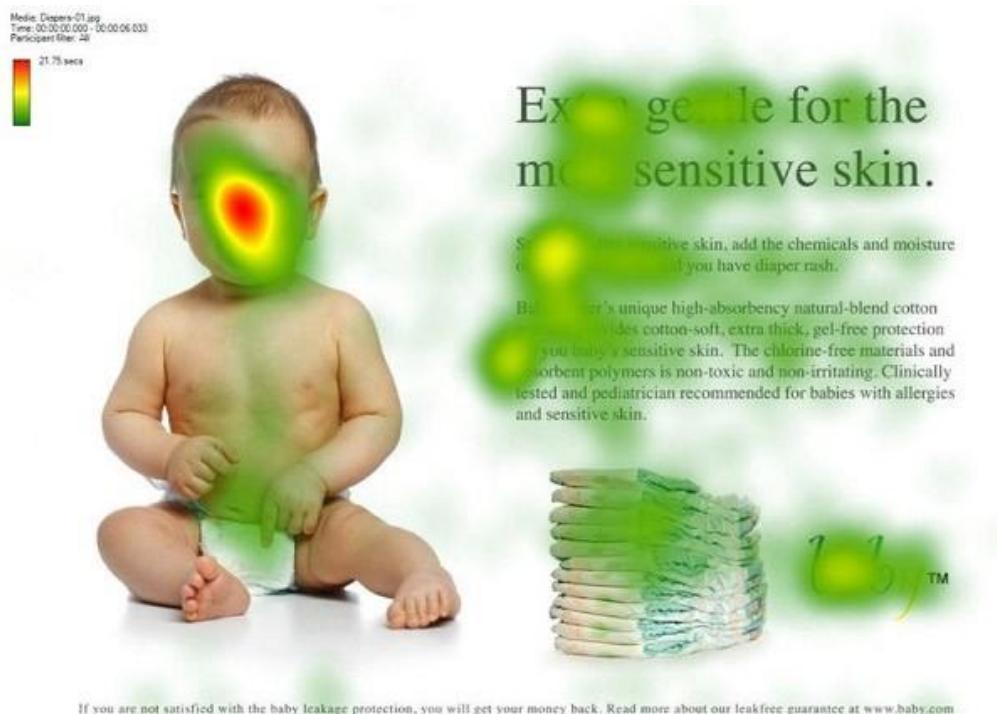


Figura 2.4 Ejemplo de mapa de saliencia (Fuente: blog de Jessica Brillhart [25])

Se han realizado experimentos que arrojan luz sobre cómo la cognición humana funciona en realidad virtual. En concreto, Serrano et al. [15] investigaron la forma en que el ser humano realiza la segmentación de eventos en este nuevo entorno: si ésta ocurre de igual forma que en el cine convencional, y cómo se desenvuelve la atención de los espectadores en los cortes. En su experimento, varias decenas de sujetos fueron sometidas a dos pruebas: la primera, replicando el procedimiento del experimento que Zacks y Magliano [17] habían realizado con cinematografía tradicional, consistió en la visualización de cuatro cortometrajes en RV, durante los cuales tenían que identificar los límites de los eventos ocurridos en la historia. La segunda prueba expuso a los sujetos una serie de clips de 360° de 12 segundos, a mitad de los cuales ocurría un corte cinematográfico; se realizó un seguimiento de la pupila para estudiar la saliencia y analizar varios parámetros característicos del comportamiento del espectador tras el corte. Los resultados obtenidos mostraron similitudes con los de Zacks et al. [19] y de Zacks y Magliano [17], dado de que el número de fronteras de evento que se detectó coincidiendo con un corte en continuidad fue significativamente bajo, demostrando que los cortes de este tipo en RV mantienen la sensación de fluidez narrativa.

## 2.3 Sonido panoramizado con orientación, o *panning*

Se ha visto qué ocurre en los cortes de edición cinematográfica en RV; Serrano et al [26] estudiaron algunos casos en este campo, pero utilizaron sonido en estéreo sin distribución panorámica en función de la orientación. Esta técnica, conocida como *panning*, recibe su

nombre de *pan*, la abreviatura de *panoramic*, haciendo referencia a la panoramización del sonido de un contenido audiovisual. El panning equivale a la distribución de la señal sonora, ya sea monoaural o estereofónica (lo que comúnmente conocemos como *mono* y *estéreo*), en un nuevo medio estéreo o multicanal controlado por un potenciómetro panorámico (o *pan pot*) que la desvía a uno u otro canal en función de las necesidades del editor. En otras palabras, permite distribuir la señal sonora con mayor intensidad a un altavoz que a otro, pudiendo crear la sensación de que el sonido proviene particularmente de una dirección concreta, como se puede apreciar en la Figura 2.5. En el caso de una disposición de altavoces o auriculares en estéreo, distribuyendo la señal enviada a ambos dispositivos y realizando la triangulación adecuada se puede simular un foco de sonido en cualquier dirección a nuestro alrededor.

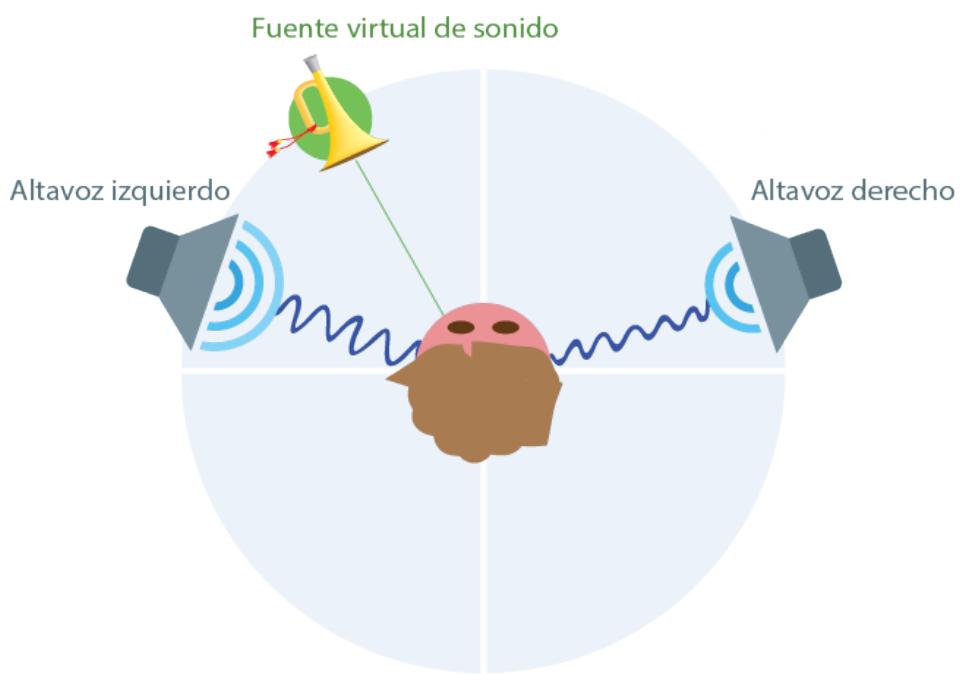


Figura 2.5 Funcionamiento del panning orientativo: la mayor intensidad del altavoz izquierdo hace interpretar al sujeto que la fuente de sonido se encuentra hacia ese lado.

Por ello, el panning es una herramienta sumamente útil en realidad virtual, dado que confiere un grado mayor derealismo a la escena. Permite al usuario, equipado con auriculares, percibir que el sonido llega desde donde virtualmente ha sido producido en el entorno en el que está inmerso, gracias a la adecuada composición geométrica de la intensidad de la señal sonora en cada auricular y a la habilidad innata del ser humano de localizar la dirección en la que se encuentra la fuente por triangulación automática, como se puede ver en la Figura 2.6. Para realizar correctamente dicha distribución de la señal se requiere un software específico que incluye la función de detección de la orientación del espectador, para, conociendo el punto de procedencia virtual (foco virtual en la escena) del sonido, calcular la intensidad de la señal de audio que debe emitir cada fuente.



Figura 2.6 Localización de la fuente de un sonido mediante triangulación.

Dado que los experimentos realizados por Serrano et al. [26] no incluyen esta distribución panorámica del sonido, se estima necesario realizar una investigación sobre el efecto de dicha técnica en el espectador en los cortes cinematográficos.

## 2.4 Edición cinematográfica: J-cut

Uno de los recursos sonoros frecuentemente empleados en el cine es el denominado J-cut. Este tipo de corte cinematográfico consiste en el desajuste de la imagen y el sonido brevemente antes de un cambio de escena, pudiéndose escuchar unos pocos segundos el audio de la nueva toma mientras aún se está visualizando la anterior. La figura 2.7 ejemplifica la edición de un J-cut en Adobe Premiere.

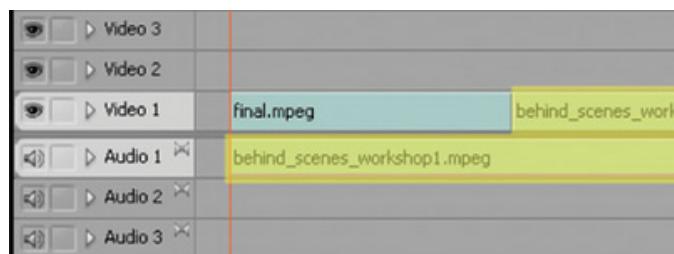


Figura 2.7 Esquema de un J-cut durante la edición. Se puede ver cómo el clip de audio correspondiente a la segunda toma entra antes que el estímulo visual. (Fuente: VideoMaker.com [27])

Existe un corte gemelo, llamado L-cut, que consiste en la prolongación del audio de la primera escena respecto del corte visual en lugar de adelantarla. Su uso más habitual es un cambio de escena manteniendo el sonido debido a que las localizaciones de ambas escenas son cercanas. Suele ser más efectivo en diálogos o gritos.

Las aplicaciones que se le dan al J-cut son múltiples: puede utilizarse con el fin de preparar a la audiencia para la nueva escena antes de que esta se muestre; para mantener un ritmo determinado; para transmitir una sensación -agobio, desconcierto, desorientación-, o con intenciones meramente artísticas, como hacer coincidir temporalmente un sonido con un gesto al que no se corresponde. Como ejemplo, según nos muestra la Figura 2.8, encontramos un J-cut a partir del segundo 00:22 del corto Waltz for One (Intellectual Propaganda, 2012) en el que el sonido del astronauta se puede oír antes de aparecer éste (vídeo disponible: [28]).



(1)



(2)

Figura 2.8 Ejemplo de J-cut (ir a [28] para acceder al enlace). En primer lugar observamos la cápsula flotando en el espacio (1) (00:24 seg), y empezamos a oír los ruidos de la radio y de la maquinaria de la nave, lo cual sería imposible dado que en el espacio no se propaga el sonido. Luego ocurre el corte visual, y aparece el casco del astronauta (2) (00:35 seg), escena a la que corresponde el sonido ya oído.

## 2.5 Motivación e hipótesis

Como se ha visto, el audio con panning direccional puede suponer un cambio drástico en la interacción con el usuario en realidad virtual. Dado que los experimentos realizados por Serrano et al. [15] no se han realizado con esta técnica y se desconocen sus efectos, se cree indispensable hacerla el pilar central de nuestro estudio. Según nuestra hipótesis, su uso reducirá el tiempo que tarda el espectador en encontrar a los personajes, así como contribuirá a minimizar el error en el camino seguido por los ojos en la búsqueda de una ROI.

Respecto al J-cut, dada su demostrada capacidad de facilitar la introducción de la nueva escena al espectador en la cinematografía tradicional, pensamos que puede resultar una herramienta útil a la hora de controlar la atención del observador de realidad virtual; es por ello que será objeto de estudio en uno de los experimentos de este trabajo. Creemos que puede avisar al usuario del cambio de escena que está por llegar, y más importante aún, siendo además capaz de anticiparle la nueva localización de los personajes o de la acción relevante.

## 3. Experimentos

---

Tras el estudio del estado del arte de la cinematografía en realidad virtual se busca comprender mejor el comportamiento del usuario de RV frente a un contenido cuyo fin es contar una historia, es decir, cómo se comunica la continuidad narrativa al espectador valiéndose de la edición y el montaje audiovisual; siendo el sonido la principal protagonista de nuestra observación. Para ello, conociendo los estudios en la materia realizados por Serrano et al. [15], se plantean dos experimentos centrados en constatar la influencia de técnicas sonoras en el proceso comunicativo: el primero consiste en estudiar el efecto del uso de audio con distribución panorámica o panning; y el segundo, de concepto y realización similares al primero, busca analizar la influencia del uso de la técnica de edición J-cut en la percepción del usuario.

El carácter de los experimentos requiere la participación de un gran número de personas, pues como sostiene Sebastian Sylwan [24]: “*Para establecer un lenguaje se requiere ver la reacción del mayor número posible de personas. [...] Cuanta más gente tengamos [consumiendo cine en RV] mejor podremos definir qué es lo que funciona y lo que no*”. En el primer y segundo experimentos participan 35 y 26 personas, respectivamente.

Nota: Para facilitar la distinción entre el trabajo realizado por el autor de esta memoria y los realizados previamente por otros autores en los que se basa, durante la narración de los experimentos se emplea el tiempo verbal presente para describir el primero, y tiempos verbales en pasado para referirse a los últimos.

### 3.1 Conceptos involucrados

Para trabajar con los parámetros y las variables es necesario definir primero una serie de conceptos, algunos de los cuales ya han sido introducidos:

### 3. Experimentos

---

- *ROI (region of interest)*: región de interés. Es la zona de la imagen de 360º donde se encuentra la información más relevante para la historia. Es por tanto la región que el creador quiere que sea vista por el espectador.
- *FOV (field of view)*: campo de visión. Es el área total visualizada por el individuo en cada instante. Comprende un ángulo sólido constante que coincide con el tamaño de la pantalla. Lo ideal es que el FOV contenga la ROI en todo momento; esto significará que el usuario está dirigiéndose a donde se encuentra contenida la información más importante.
- *Estímulo*: es el contenido mostrado al sujeto del experimento. Debe cumplir una serie de requisitos que dependen del experimento a realizar y sus condiciones.
- *Saliente*: es la capacidad de un elemento (estímulo en nuestro caso) para llamar la atención del observador por encima del entorno. Un mapa de saliente representa, generalmente en forma de mapa de calor, la saliente de cada punto de una imagen.
- *Fijación y sacada*: respectivamente, son la posición estática de la mirada en un punto, y el movimiento rápido del globo ocular entre una fijación y la siguiente. La conjunción de ambas conforma el seguimiento ocular de una imagen.
- *Vídeo 360º*: contenido audiovisual en formato para ser visualizado en realidad virtual. Se visualiza proyectando las imágenes sobre la superficie interior de una esfera, desde el centro de la misma y abarcando casi la totalidad de su superficie.

El sistema de visión del ser humano no es capaz de asimilar toda la información del campo visual disponible; más del 90% se percibe con baja resolución. Esto es así debido a que la mayor parte de los fotorreceptores se encuentran en la *mácula*, especialmente en la llamada *fóvea* de la retina, lugar donde se proyecta el centro de la imagen que atraviesa la pupila, ligeramente por encima del eje óptico (ver Figura 3.1). Por tanto, para generar una imagen completa del campo de visión en nuestro cerebro necesitamos fijar la mirada en distintos puntos, entre los cuales el globo ocular se moverá con sacadas. Esto algo que se realiza de forma automática, mostrando el ser humano un comportamiento exploratorio a la hora de analizar una imagen ([29, 30, 31]).

Durante la visualización de un vídeo 360º en realidad virtual puede ocurrir que, aunque la ROI esté dentro del FOV, la fijación no coincide dentro de la *ROI*. Sin embargo, debido a este comportamiento exploratorio natural del ser humano y a que la ROI tiene la saliente más alta supuestamente -si la creación del contenido se ha realizado correctamente-, lo más probable es que el espectador acabe mirando a esta región que posee la información más relevante.

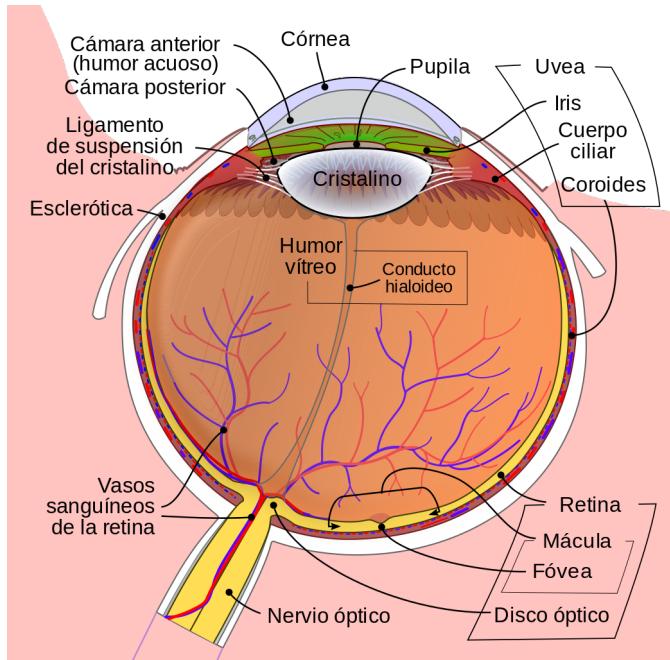


Figura 3.1 Diagrama del ojo humano (Fuente: Wikimedia Commons [32])

## 3.2 Primer experimento: sonido con *panning* orientativo

El objetivo de este experimento es cuantificar el efecto causado por el uso de audio con panning orientativo en la percepción de la continuidad narrativa en los cortes cinematográficos en realidad virtual. Dado que conocemos los detalles de un experimento similar realizado con éxito -la investigación de Serrano et al. [26]- y resulta interesante poder comparar nuestros resultados con aquél, utilizaremos sus mismas condiciones y estímulos, replicando el experimento salvo por la inclusión del sonido distribuido panorámicamente en función de la dirección de observación. Estas condiciones se explican a continuación.

### 3.2.1 Planteamiento

El experimento consiste en exponer un estímulo, en este caso un vídeo 360º en RV formado por dos escenas breves separadas por un corte cinematográfico, a un conjunto de individuos en condiciones controladas y de forma individual, monitorizando ciertos parámetros, y midiendo la saliencia de las imágenes antes y después del corte.

Las variables a estudiar son escogidas de forma que representen lo mejor posible el amplio abanico de escenarios. Coinciendo con los utilizados por Serrano et al. [26], se han considerado:

1. El desalineamiento de las ROIs antes y después del corte, A, medido en grados horizontales desde el centro de cada ROI en el vídeo 360º.

### 3. Experimentos

---

2. La configuración de las ROIs antes y después del corte, R, que incluye su número y posición.

3. El tipo de corte cinematográfico, E, pudiendo ser de tres tipos: con continuidad en acción, en espacio/tiempo, o ambos.

#### 3.2.2 Preparación

Los estímulos utilizados se escogen de entre los clips de vídeo usados por Serrano et al. [15]. Su preparación se realizó a mano con el programa Adobe Premiere Pro CC 2017 por otros integrantes del laboratorio. En primer lugar se grabaron distintas tomas de cuatro escenas cotidianas -trabajo realizado por la empresa dedicada a la grabación de vídeos 360° Abaco Digital-, a partir de las cuales se montaron los vídeos con la siguiente estructura: dos escenas de 6 segundos separadas por un corte cinematográfico convencional, sumando así clips de 12 segundos de duración. La resolución de los vídeos es de 3840x1920 píxeles y su tasa de fotogramas por segundo es muy próxima a 60 fps.

Las escenas se grabaron en cuatro localizaciones distintas, como se muestra en la Figura 3.2: una cocina, unas escaleras, un aula y un salón. En ellas las tomas se realizan de 3 formas distintas según la configuración de las ROIs. Estas regiones son sujetos realizando tareas simples y fácilmente identificables: los actores -trabajadores del laboratorio- llevan a cabo acciones cotidianas como fregar, recoger la vajilla, subir las escaleras, leer un libro, consultar el móvil, escribir en un papel o preparar un café.



Figura 3.2 Las cuatro localizaciones de las escenas: cocina (arriba, izquierda), escaleras (arriba, derecha), aula (abajo, izquierda) y salón (abajo, derecha)

Los desalineamientos deben muestrear distintas situaciones características de la visualización de contenido en realidad virtual, por lo que los utilizados en el experimento son los siguientes:

- $A_{40}$ :  $40^\circ$  de separación entre la orientación inicial y la ROI en el eje horizontal
- $A_{80}$ :  $80^\circ$  de separación entre la orientación inicial y la ROI en el eje horizontal

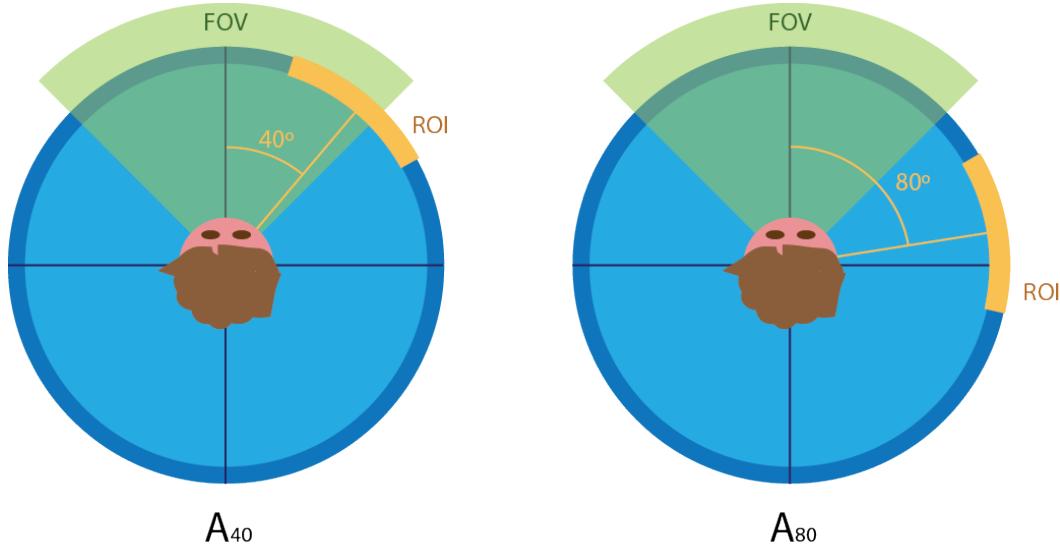


Figura 3.3 Esquema de los dos desalineamientos

Se eligen estas cantidades angulares teniendo en cuenta que una ROI abarca aproximadamente unos  $40^\circ$  en horizontal, mientras que el FOV, por las especificaciones técnicas del casco HMD -ver Capítulo 4, Sección 4.1- ocupa  $95^\circ$ . Así, se garantiza que en el primer caso,  $A_{40}$ , la ROI queda parcialmente dentro del FOV y por tanto a la vista del sujeto, y en el segundo caso,  $A_{80}$ , queda fuera de ésta, como puede apreciarse en la Figura 3.3.

Las distintas configuraciones de las ROIs responden a 3 posibles distribuciones antes del corte (ver Figura 3.4), llamadas  $R_b$  (*ROIs before*), combinadas con otras 3 posibles después del mismo, llamadas  $R_a$  (*ROIs after*), lo que suma un total de 9, como se muestra en la Figura 3.5. Las 3 distribuciones son:

- Una ROI, equivale a la variable  $R_{a|b}=0$
- Dos ROIs en el mismo FOV (se pueden ver los dos a la vez), equivale a  $R_{a|b}=1$
- Dos ROIs en distinto FOV (no se pueden ver los dos a la vez), equivale a  $R_{a|b}=2$

### 3. Experimentos

---



Figura 3.4 Configuraciones de las ROIs

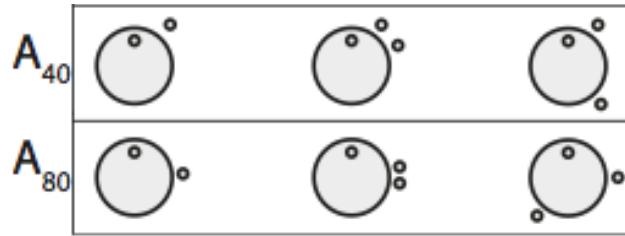


Figura 3.5 Combinaciones de las ROIs con los desalineamientos

Para reducir el sesgo derivado de las particularidades de cada escena individual, se determina utilizar 4 distintas versiones de cada perfil de clip -pero con la misma configuración-. Dados los 3 tipos de cortes, los 2 desalineamientos y las 9 posibles configuraciones, se necesitan  $3 \times 2 \times 9 \times 4 = 216$  clips distintos para que todas las combinaciones de los estímulos tengan su muestra en el experimento.

Sin embargo, a la hora de seleccionar los clips de vídeo de entre los disponibles, se tienen en cuenta una serie de restricciones destinadas a minimizar los factores sonoros que puedan interferir en la correcta realización de las pruebas, y a reducir la aparición de sesgos:

- El sonido de cada personaje debe tener similar intensidad
- La lejanía de las distintas fuentes de sonido ha de ser parecida
- Todas las configuraciones de desalineamientos y ROIs antes y después del corte deben estar presentes en el experimento
- La fuente virtual del sonido debe poder seguir a los personajes en la escena con un error menor a la anchura angular de una ROI

Ciertas complicaciones surgen durante la preparación del experimento: debido a que no se dispone de suficientes clips que cumplan estas restricciones, se elimina del estudio la variable referente al tipo de corte, E, debido a que es la que menor variación demuestra provocar en los sujetos de los experimentos de Serrano et al [26], y no resulta interesante para nuestro estudio. Por otro lado, de cada perfil de clip se seleccionan 3 variantes de escenas distintas

para reducir los efectos de cada escena en particular, en lugar de 4. Se descarta por tanto una parte de los clips que se habían empleado en el experimento original de Serrano et al.

Finalmente resulta un total de  $2_{\text{desalin.}} \times 9_{\text{config.}} \times 3_{\text{clips}} = 54$  clips, en cuyas variaciones incluyen todos los desalineamientos de las ROIs así como sus posibles configuraciones antes y después del corte.

Para visualizar los vídeos 360° adecuadamente se utiliza un casco HMD, modelo Oculus DK2. Asimismo, se necesita un software que reproduzca el contenido a mostrar en el casco, para lo cual se emplea Unity3D, un software de desarrollo de videojuegos y entornos virtuales que se puede controlar mediante ficheros (o *scripts*) de código en el lenguaje de programación C#; además, este motor gráfico se encarga de reproducir las pistas de audio con panning direccional a través de los auriculares, creando la sensación de estar escuchando sonido panoramizado en el espectador.

Para conseguir que Unity3D reproduzca el sonido de forma panorámica desde donde están los personajes, se crea una función en el código en C# del programa (ver Capítulo 4.1 para mayor información) que se encarga de mover los altavoces virtuales del audio allá donde están los sujetos del vídeo en el espacio virtual, como se puede ver en la Figura 3.6:



Figura 3.6 Distribución de los altavoces virtuales siguiendo a los personajes (círculos azules sobre las cabezas de Víctor y Jaime)

### 3. Experimentos

---

Estos altavoces virtuales de Unity reproducen las pistas de audio de la prueba en función del vídeo visualizado -pistas que se editan con Adobe Premiere CC 2017. Se crea una pista para cada fuente de sonido, que coincide con los personajes -y con las ROIs- de la escena; así el programa es capaz de, conociendo las coordenadas de los personajes en el espacio virtual, panoramizar el sonido como si procediera de ellos. Estas coordenadas son calculadas a mano previamente para todos los personajes de cada clip y registradas en un fichero de texto, al que debe acceder Unity3D para colocar en la escena las fuentes virtuales del audio.

La información a estudiar del sujeto se obtiene durante la prueba mediante el seguimiento ocular con las cámaras *eyetracker* de PupilCapture instaladas dentro del casco HMD como muestra la Figura 3.7, que incluyen el software necesario para manejarlas y recoger el flujo de datos mientras se realiza el experimento.



Figura 3.7 Detalle del montaje del eyetracker (Fuente: Pupil Labs [33])

El montaje final queda compuesto por dos ordenadores, debido a que se requiere una altísima potencia computacional para hacer funcionar el entorno virtual. En un ordenador con Windows 8 (izquierda en la Figura 3.8) se instalan el motor gráfico Unity3D y el dispositivo de realidad virtual de Oculus que se emplea en la prueba, el casco Oculus DK2. En el ordenador de la derecha de la Figura 3.8 se instala PupilCapture, el software de los seguidores oculares.



Figura 3.8 Montaje del sistema para el experimento (no definitiva)

Mediante un proceso de prueba con diferentes sujetos voluntarios, entre ellos el autor, se verifica que todo funciona correctamente y se puede comenzar a realizar las pruebas de usuario.

### 3.2.3 Realización del experimento

Los sujetos de las pruebas realizan una breve encuesta inicial acerca de su salud ocular y auditiva. Seguidamente se les coloca el casco HMD Oculus DK2 y los auriculares de audio de alto aislamiento acústico Vic Firth. Tras ajustar los dispositivo correctamente, comienza la prueba con la calibración de los *eyetrackers*, la cual permitirá a estas herramientas tomar las medidas correctas de la posición de la mirada. Una vez terminada la calibración empieza la sucesión de vídeos 360° de 12 segundos seleccionados al azar, hasta ver todas las configuraciones de ROIs una sola vez. Si un sujeto ve un clip desalineado A<sub>40</sub>, no se le mostrará el mismo vídeo desalineado A<sub>80</sub>, y viceversa, para evitar sesgar su comportamiento mientras ve el segundo. Esto significa que cada persona verá 27 vídeos, tendiendo a ser, estadísticamente, la mitad con desalineamiento A<sub>40</sub> y la otra mitad con A<sub>80</sub> de forma aleatoria.

Mientras el sujeto mira los vídeos y sigue las acciones llevadas a cabo por los personajes, los *eyetrackers* registrarán la posición de la pupila y se tomará también la orientación del casco HMD, pudiendo calcular posteriormente la localización exacta de las fijaciones oculares en el video y así conocer su recorrido por la escena.

Finalmente, los participantes contestan a unas preguntas referentes a la experiencia que han vivido: las sensaciones que han sentido, si les ha resultado cómodo, etc. Para ver la encuesta inicial y final se puede consultar el Anexo D. 35 personas realizaron este experimento.

### 3. Experimentos

---

#### 3.2 Segundo experimento: J-cut

Con este experimento se pretende comprobar si el J-cut representa una mejora en la percepción de la continuidad narrativa en los cortes cinematográficos en realidad virtual. Esta prueba es similar al primer experimento; mantiene el procedimiento, la toma de datos, las condiciones y configuración de aquél con una simple pero importante diferencia: el corte en los vídeos no es uno convencional, sino que se utiliza el J-cut. Como se ha explicado anteriormente, este corte permite desajustar el momento del cambio de escena del sonido respecto de la imagen, ocurriendo el del primero unos segundos antes que el segundo. Así, en este experimento se alternan cortes de 1 y de 2 segundos de adelanto del sonido con respecto al cambio de escena.

Otra diferencia con respecto al primer experimento, no por planteamiento sino por inconveniencia, es que durante la preparación de los estímulos, 10 de los clips de 12 segundos no han podido ser adaptados al J-cut debido a que el fragmento de 6 segundos posterior al corte se tomó desde el inicio de la escena grabada, por lo que no se ha disponido del audio desde 1 o 2 segundos antes. Esto significa que de los 54 vídeos, en este experimento los sujetos ven solamente 44, siendo la mitad con un J-cut de 1 segundo y la otra mitad de 2 segundos.

De nuevo, en este experimento los sujetos ven cada clip una sola vez, siendo estos mostrados de forma aleatoria, y evitando así el sesgo de la visualización *a posteriori*.

Tras terminar el experimento, se cuenta la participación de 26 sujetos.

Para terminar, una vez dados los dos tests por finalizados, se sabe que un total de 61 personas acudieron a participar

## 4. Entornos de trabajo

---

Los experimentos realizados en este trabajo esconden detrás un completo entramado de dispositivos, muchos de los cuales ya han sido mencionados. Estos dispositivos son los encargados, por una parte, de hacer funcionar en el casco HMD y los auriculares el mundo virtual deseado, y por otra, de la adquisición de datos que serán analizados y nos servirán finalmente para encontrar las respuestas a nuestras preguntas.

### 4.1 Entorno audiovisual

#### 4.1.1 Oculus Development Kit 2

El casco HMD de realidad virtual utilizado en los experimentos es de la marca Oculus VR, siendo el segundo de los kits de desarrollo del conocido Oculus Rift.

Este casco está formado por una pantalla del tamaño aproximado de un smartphone (unas 5,7 pulgadas) y una resolución de 1920x1080 píxeles en formato 16:9 dentro de una carcasa negra, como se muestra en la Figura 4.1. Esta carcasa, abierta por un lado, permite al usuario mirar la pantalla interna a través de unas lentes que dirigen sus rayos de luz hacia nuestros ojos como si vinieran del infinito -puesto que las lentes se colocan de forma que la retina ocular se encuentre en su plano focal-, y protege los ojos de la luz exterior, de forma que lo que se percibe son únicamente sendas imágenes en cada ojo. Por tanto, las dimensiones de la pantalla se dividen entre dos, quedando un formato 8:9. Tal es la configuración del sistema, que aporta un FOV de 95° horizontales por 106° verticales.

El HMD de Oculus incorpora sensores de rotación (giroscopio), aceleración y una brújula que le permite conocer la orientación absoluta, responsables del registro de la dirección de la cabeza durante los experimentos.

## 4. Entornos de trabajo

---



Figura 4.1 Detalles del casco HMD de Oculus (Fuente: Oculus [3])

### 4.1.2 Unity3D

Para proyectar el mundo virtual en un casco HMD como el de Oculus, es necesario un software que funcione como motor gráfico. Unity3D no sólo es interesante por su versatilidad a la hora de generar contenido de videojuegos y de realidad virtual, sino por ofrecer una interfaz relativamente simple para los no iniciados en la programación a la vez que permitir un avanzado control y desarrollo de los proyectos para los más doctos en la materia.

Un proyecto de Unity trabaja con escenas, dentro de las cuales existe una jerarquía de objetos con sus propiedades y sus funciones. Estos objetos se pueden crear y se les puede asignar sus cualidades, o también se pueden construir a partir de predefinidos existentes en las librerías de Unity. Las escenas simulan el espacio tridimensional real con los tres ejes coordinados, como muestra la Figura 4.2, y los objetos ocupan una posición en él.

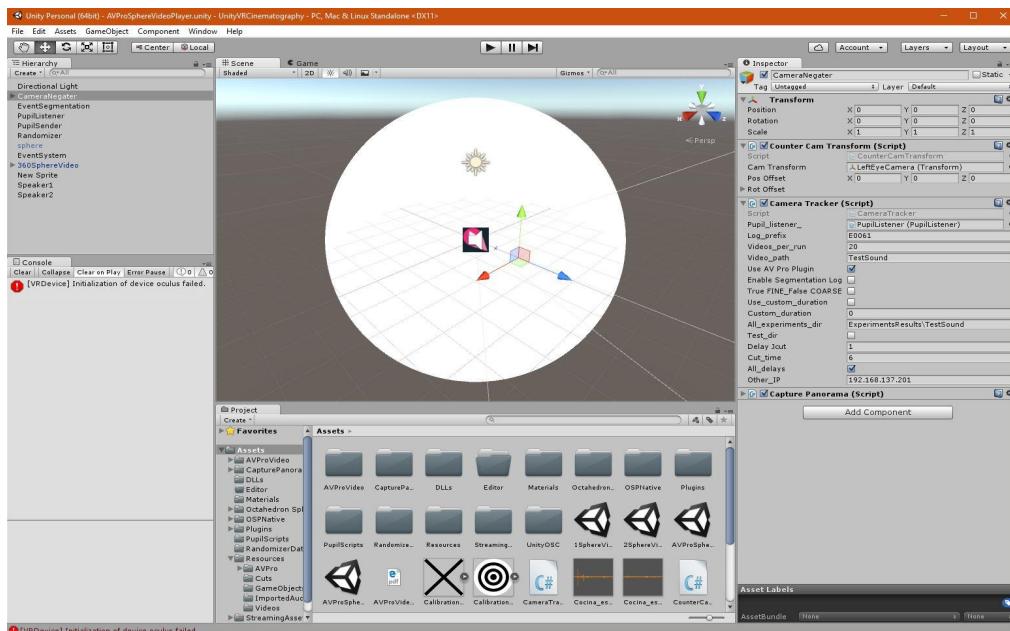


Figura 4.2 Interfaz de Unity3D en Windows 8

En Unity se puede programar el comportamiento de los objetos mediante código de programación. En este trabajo se utilizó el lenguaje C# específicamente por estar orientado a objetos, particularidad que encaja muy bien con la filosofía de Unity.

La escena utilizada en este programa está compuesta por los siguientes objetos: el reproductor de vídeo; el emisor que proyecta sobre una esfera (ver Figura 4.3) la imagen del vídeo 360° que se va a ver alrededor del espectador, siendo la esfera a su vez también un objeto; el foco que ilumina la escena; los altavoces que reproducen las pistas de audio desde su localización en el espacio tridimensional; las cámaras que siguen los ojos; y el *camera tracker* o controlador encargado de coordinar todas las funciones de los demás objetos. Todos están controlados por líneas de código ordenadas en ficheros o *scripts* asociados a cada objeto, que permiten conferirles un comportamiento determinado y encomendar la ejecución de tareas al motor gráfico.

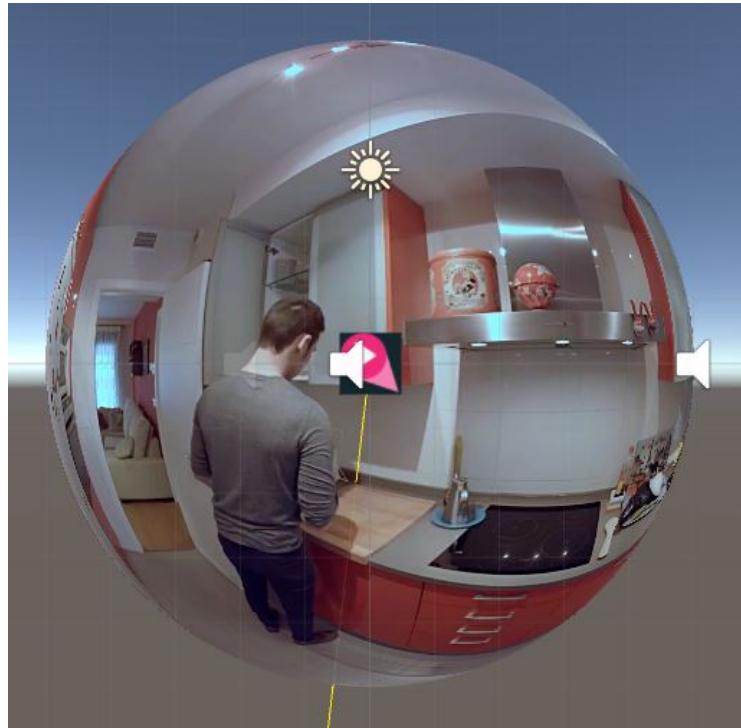


Figura 4.3 Objeto esfera donde se proyecta la imagen 360°

Los objetos con función de reproductor, proyector y altavoz son los responsables de tomar los archivos de contenido -los clips y los audios- y mostrarlos a través de la pantalla del Oculus DK2. El *camera tracker* será el encargado de acceder a la información obtenida por el código que permite escoger los vídeos de forma aleatoria (*randomizador*), asignar la posición de los altavoces virtuales en la escena de Unity en función de la localización de los focos de sonido en los clips (personajes), e intervenir en la comunicación con otros dispositivos externos al programa como PupilCapture, entre otras funciones.

## 4. Entornos de trabajo

### 4.1.3 Adobe Premiere CC2016

Este software ha sido utilizado principalmente para la creación de los clips, además de para definir las regiones de interés de las escenas y a partir de ellas sacar las métricas. Ha permitido la edición separada de los videos y los audios, resultando el preparar los J-cuts una tarea sencilla. La Figura 4.4 muestra un ejemplo del interfaz.



Figura 4.4 Interfaz de Adobe Premiere CC 2016

## 4.2 Entorno de adquisición de datos

Además del registro en tiempo real de la dirección de la cabeza del sujeto durante el experimento, que es llevado a cabo por el casco HMD de Oculus DK2, el principal componente de adquisición de información es PupilCapture.

### 4.2.1 PupilCapture 0.8.7

Es a la vez el hardware y el programa que se encarga de registrar en cada fotograma del clip mostrado el punto de la imagen al que está mirando el espectador durante un experimento. Lo consigue gracias a dos cámaras, una para cada ojo, que realizan un seguimiento continuo de la pupila desde detrás de las lentes del casco HMD de Oculus.

PupilCapture pertenece a Pupil Labs, una empresa especializada en el desarrollo de software y hardware que gira en torno al seguimiento de las pupilas. Pupil Labs trabaja con código libre, por lo que es fácilmente adaptable y ampliable a otras áreas de trabajo.



Figura 4.5 Detalles del montaje del *eyetracker* en el Oculus DK2 (Fuente: Pupil Labs [33])

El montaje se realiza colocando los detectores tras las lentes, como se puede observar en las Figuras 4.5. Sendos cables se conectan con el ordenador encargado de recibir la información, un Linux con Ubuntu en nuestro caso, que tiene instalado el software de control de PupilCapture; en la Figura 4.7 se puede observar el seguimiento que realiza de la pupila.

## 4. Entornos de trabajo

---

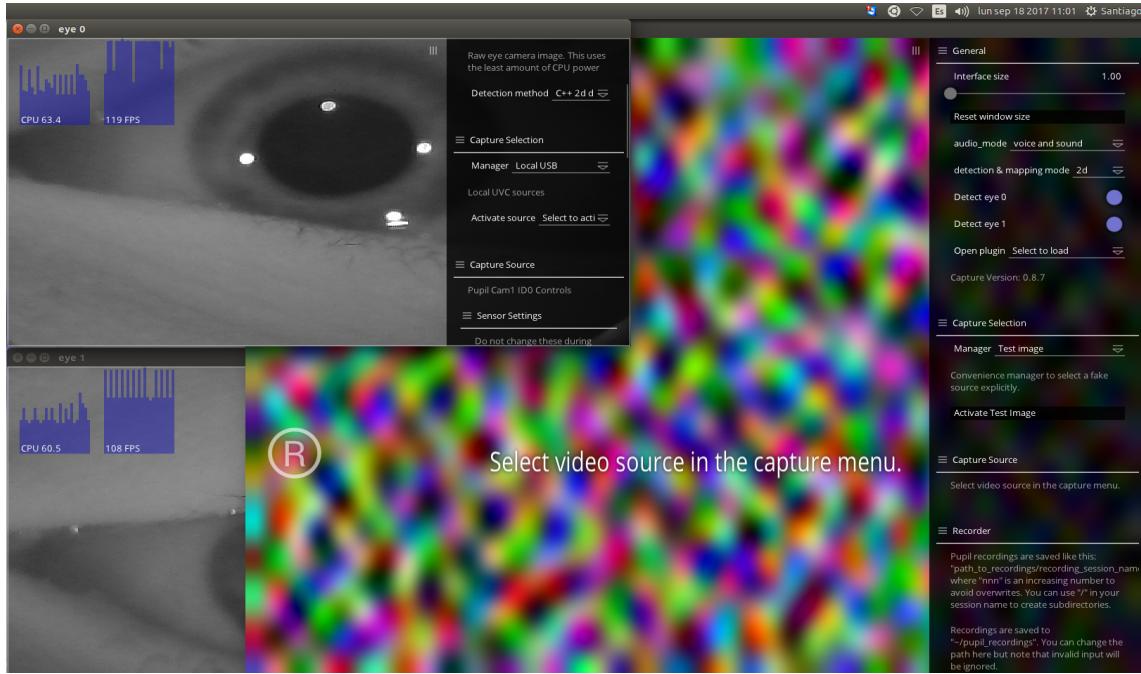


Figura 4.6 Ojo del sujeto en la interfaz del programa PupilCapture en Ubuntu

Una vez instalado el programa y las cámaras, se configura para ser utilizado en realidad virtual. Cada vez que se quiera utilizar el *eyetracker* será necesario calibrar los detectores; en este caso esto se conseguirá mostrando una X en la pantalla del casco HMD que se desplazará a determinadas posiciones, haciendo a los ojos del sujeto seguir su recorrido. De esta forma el eyetracker sabrá a qué punto de la imagen corresponde cada posición de las pupilas, y podrá guardar esta información para ser analizada posteriormente.

Desde el inicio del experimento, los seguidores oculares registran la posición de las pupilas a un ritmo superior a 120 fps (fotogramas por segundo), más del doble que el ritmo de fotogramas por segundo de los vídeos.

## 5. Análisis de datos

---

Tras la recopilación de los datos de los experimentos son necesarios un tratamiento y posterior análisis de la información recogida. En primer lugar se realiza un pre-procesado, que ordena y prepara los datos para ser tratados de forma óptima. Posteriormente se obtienen las métricas (o medidas) a partir de los parámetros que se han monitorizado; éstas nos permitirán conocer de forma cuantitativa el comportamiento de los sujetos frente a los estímulos mostrados. Con estas medidas seremos capaces de valorar los resultados y finalmente elaborar las conclusiones.

Este es un proceso muy técnico tanto de llevar a cabo como de narrar, por lo que se ha resuelto exponerlo en el anexo en lugar de en la memoria. Para mayor detalle y profundización en el procedimiento, se ruega consultar el Anexo C.

### 5.1 Estudio de las métricas

Las métricas utilizadas en estos experimentos son análogas a las empleadas por Serrano et al [26], debido a que este proceder ofrece la posibilidad de comparar ambos resultados. Cuando se planteó el experimento de Serrano et al. se buscaba analizar el comportamiento del espectador de cinematografía en RV frente a los cortes, por lo que se diseñaron unas medidas específicas que representaban características propias de la conducta de un sujeto al visualizar un contenido audiovisual. Estas son:

- *Número de fijaciones*: representa el ratio de fijaciones con respecto a la totalidad de las muestras de mirada tomadas por el eyetracker después del corte cinematográfico, una vez encontrada una ROI. En otras palabras, compara cuántas fijaciones y sacadas hay en el barrido ocular de la imagen, sugiriendo un comportamiento más o menos exploratorio; cuanto menor sea el valor de esta métrica, menos fijaciones se habrán producido en una zona de la imagen, dando a entender que la saliencia de esta área es baja.

## 5. Análisis de datos

---

- *Fotogramas en alcanzar una ROI:* contabiliza el número de fotogramas que le cuesta al observador encontrar una ROI después del corte cinematográfico. Dado que los fotogramas se actualizan a una frecuencia determinada (60 fps), y la frecuencia se define como eventos/tiempo (#/segundo en el Sistema Internacional), esta métrica es equivalente a medir el tiempo que tarda la mirada en apuntar dentro de una ROI.
- *Porcentaje de fijaciones dentro de la ROI:* cuando después de un corte el sujeto encuentra una ROI, se produce la primera fijación dentro de ésta. A partir de ese momento, esta métrica contabiliza cuántas fijaciones tras el corte caen dentro de la zona de interés, y las compara al total de fijaciones después del corte. Esta medida nos indica el grado de interés del observador en las ROIs.
- *Error en la trayectoria en la búsqueda de ROIs:* se calcula la raíz cuadrada del error cuadrático medio (RMSE) para la diferencia entre la trayectoria seguida por la mirada del sujeto y la trayectoria media del baseline. Para hacerla independiente de Fotogramas en alcanzar una ROI, se calcula a partir del momento en el que se produce la primera fijación en una ROI tras el corte.

### 5.2 Análisis

Como se explica en el Anexo C, una vez introducida en el programa de análisis de datos y estadística *SPSS17* la información procesada resultante de los experimentos, se calculan los valores medios de las métricas, su desviación estándar y se someten a un análisis *Multilevel Modeling* [34] de efectos combinados (*Mixed Model Analysis*), el cual nos indica, mediante el *valor de significancia* de los parámetros, de qué forma las métricas se ven afectadas por cada una de las variables, y en qué medida. También nos aporta información acerca de su dependencia del sujeto, según la significancia de la variancia de la variable que identifica a cada elemento de la muestra.

Las métricas estudiadas demuestran ser dependientes del sujeto, por lo que no es posible un análisis *ANOVA* de los datos. En el análisis se comprueba la influencia de las tres variables estudiadas: el deslineamiento de las ROIs antes y después del corte, las configuraciones de los ROIs –tanto antes como después del corte–, y el J-cut.

## 6. Resultados

---

Aquí se muestra el fruto del procesado y análisis de la información recopilada durante los experimentos estudiando el impacto de las variables en los parámetros medidos.

### 6.1 Primer experimento

#### 6.1.1 Fotogramas en alzanzar una *ROI*

Queremos cuantificar la variación entre en la respuesta de los sujetos frente a nuestros experimentos con audio panorámico y los estudiados por Serrano et al [26]. Para ello, en primer lugar, en el capítulo anterior se ha sometido la métrica *fotogramas en alcanzar una ROI* al *Mixed Model Analysis* de SPSS, cuyo cálculo de la significancia (ver Anexo C.3) demuestra que la variable del desalineamiento es significativa, es decir, supone una diferencia sustancial en el parámetro a estudiar, al tener un valor inferior a 0,05. También la direccionalidad del sonido con panning resulta ser significativa. Por tanto, nos interesa comparar el valor medio de los *fotogramas en alcanzar una ROI* en función de ambas variables de estudio, habiendo calculado con SPSS los resultados de la Tabla 6.1:

<i>A, desalto.</i>	<i>Nuestra media (con panning)</i>	<i>Media de Serrano (sin panning)</i>	<i>Relación</i>
$A_{40}$	$23,28 \pm 2,39$	$31,57 \pm 3,40$	-26,26%
$A_{80}$	$48,36 \pm 6,24$	$69,39 \pm 5,48$	-30,31%

Tabla 6.1 Comparativa de los resultados de la métrica *fotogramas en alcanzar una ROI*. Se enfrentan el desalineamiento A y el uso de panning orientativo.

En el desalineamiento de 40º se obtiene un decrecimiento del 26%, y del 30% para 80º. Esto demuestra que el sonido con *panning* direccional contribuye positivamente a encontrar las *ROIs*. Además, estos resultados confirman la dependencia directa de este parámetro con el

## 6. Resultados

---

ángulo de desalineamiento, aunque con una pendiente menos inclinada que sin panning orientativo.

Podemos comparar también este parámetro en función de la configuración  $R_a$  de las *ROIs*:

<b><i>R</i> después del corte</b>	<b>Nuestra media (con panning)</b>	<b>Media de Serrano (sin panning)</b>	<b>Relación</b>
0	38,56	45,20	-14,69%
1	33,88	38,62	-12,27%
2	34,71	32,50	+6,80%

Tabla 6.2 Comparativa de los resultados de la métrica *Fotogramas en alcanzar una ROI*. Se enfrentan la configuración de las ROIs y el uso de panning orientativo.

En general, como se muestra en la Tabla 6.2, vemos ligeramente reducidos los fotogramas hasta encontrar una ROI, con una excepción. Una posible explicación a esta excepción es que si no tenemos panning direccional, es más probable encontrar un ROI por casualidad independientemente de hacia qué lado giremos primero, por lo tanto disminuye la mejora que pueda introducir el sonido con panning direccional.

### 6.1.2 Error en la trayectoria en la búsqueda de ROIs

	<b><i>A</i>, desalto.</b>	<b>Nuestra media (con panning)</b>	<b>Media de Serrano (sin panning)</b>	<b>Relación</b>
<b>Antes del corte</b>	<i>A</i> 40	60,25	58,16	Similar
	<i>A</i> 80	57,48	60,40	Similar
<b>Después de la ROI</b>	<i>A</i> 40	65,97	81,56	-19,12%
	<i>A</i> 80	92,21	102,66	-10,18%

Tabla 6.3 Comparativa de los resultados de *Error en la trayectoria en la búsqueda de ROIs* frente al desalineamiento y el uso de sonido con panning orientativo

Para el comportamiento medio del sujeto antes del corte se aprecia una cierta coincidencia de los datos obtenidos con los de Serrano et al, como se ve en la Tabla 6.3

En cambio, después del corte, tras haber el sujeto encontrado una ROI, el error se ve notablemente reducido para el desalineamiento de 40º (casi un 20%), y tan sólo un 10% para el de 80º. Esto puede deberse a que el sonido con panning orientativo ayuda en la búsqueda, pero si la otra ROI está fuera del FOV, le cuesta bastante encontrarla aunque el sonido contribuya a hacerlo más rápido.

## 6.2 Segundo experimento

### 6.2.1 Fotogramas en alcanzar una ROI

El análisis multilevel determina que la variable del desalineamiento es significativa para esta métrica. Por tanto, estudiamos sus valores medios:

<b>A</b>	<b>Media</b>	<b>Media sin J-cut</b>	<b>Relación</b>
<i>A40</i>	$29,32 \pm 5,44$	$23,28 \pm 2,39$	+25,95%
<i>A80</i>	$67,14 \pm 10,10$	$48,36 \pm 6,24$	+38,83%

Tabla 6.4 Comparativa de los resultados de *Fotogramas en alcanzar una ROI* frente a su desalineamiento y a la baseline sin J-cut.

Se aprecia un contundente aumento en el número de fotogramas hasta encontrar el ROI, lo que parece demostrar que el J-cut entorpece la búsqueda de estas regiones tras el corte.

### 6.2.2 Error en la trayectoria en la búsqueda de ROIs

El análisis multilevel muestra que las variables de configuración de las ROIs supone una diferencia significativa en el parámetro a estudiar, por lo que compararemos sus medias:

	<b>A, desalto.</b>	<b>Media</b>	<b>Media sin J-cut</b>	<b>Relación</b>
<i>Antes del corte</i>	<i>A40</i>	57,04	60,25	Similar
	<i>A80</i>	59,43	57,48	Similar
<i>Después de la ROI</i>	<i>A40</i>	61,98	65,97	-6,05%
	<i>A80</i>	107,46	92,21	+16,54%

Tabla 6.5 Comparativa de los resultados de *Error en la trayectoria en la búsqueda de ROIs* frente al desalineamiento y el uso de sonido con panning orientativo

No se puede constatar ninguna diferencia remarcable antes del corte. Sin embargo, después del corte, y habiendo ya encontrado una ROI, se reduce muy ligeramente el error en la búsqueda de la ROI para 40° (cerca del 6%), y aumenta visiblemente el de 80° (+16%). Esto puede deberse a que si ambas ROIs se encuentran en el mismo FOV, y se introduce un J-cut, el sujeto reacciona cambiando rápidamente de ROI porque la puede ver

## 6. Resultados

---

instantáneamente; pero sin embargo, cuando la otra ROI está fuera del FOV, el J-cut parece actuar confundiendo al observador, haciéndole creer que ha aparecido un nuevo personaje fuera de su vista en esa escena, cuando en realidad está escuchando el sonido adelantado de una ROI de la siguiente escena tras el corte visual.

En relación al experimento con y sin J-cuts, observamos que en general esta técnica contribuye a generar confusión en el sujeto cuando éste no puede ver la ROI que está produciendo el sonido; requiere, por tanto, de unas condiciones muy específicas para ser utilizado.

## 7. Conclusiones

---

La idea de este trabajo surge al preguntarse sobre cómo contar una historia en realidad virtual. ¿Funciona idénticamente como el cine tradicional, o por el contrario necesita de la creación de un nuevo marco de referencia que dicte las pautas para hacerlo? ¿Cuáles son las normas que necesitamos desarrollar para que la comunicación y la transmisión tanto de información como de intención y emociones sea fructífera y se lleve a cabo de forma óptima?

Este trabajo, como otros en los que se basa, intenta responder a estas preguntas mediante la experimentación científica. Aborda el estudio del estado del arte en edición cinematográfica tanto tradicional como en realidad virtual, y se cuestiona cómo el ser humano percibe y segmenta los eventos y cómo funciona su atención al visualizar imágenes, con el fin de realizar una investigación experimental acerca de los efectos del uso de sonido con *panning* orientativo en la percepción y el comportamiento del observador en cinematografía en realidad virtual. Por otra parte, investiga la influencia del uso del corte *J-cut* en la percepción y la conducta del espectador.

Los experimentos toman como base el realizado entre 2016 y 2017 por Serrano et al [15], siendo ambos de tipo prueba de usuario; se replican sus condiciones y se utilizan las mismas métricas para hacer posible la comparación de ambos resultados. Diversas variables involucradas en la interacción usuario-RV se han tenido en cuenta, como la configuración de las regiones de interés o *ROIs*, o su desalineamiento angular, y se han tomado medidas cuantitativas del comportamiento del sujeto, como el número de fotogramas que tarda en encontrar una ROI.

Los resultados del primer experimento arrojan una información muy interesante, habiendo contribuido el uso del *panning* orientativo a la reducción en el tiempo de exploración del sujeto hasta encontrar una *ROI* entre un 25 y un 30%, según valores de las variables. Coincide, por tanto, con la hipótesis inicial de la que parte el estudio.

El segundo experimento, sobre el uso del *J-cut*, muestra un importante hallazgo: en general, al utilizar este tipo corte encontramos que aumenta el tiempo y el error en la búsqueda de las ROIs en comparación con el uso de *panning* orientativo sin este tipo de

## 7. Conclusiones

---

corte. Por tanto, ha lugar a considerar que el J-cut, utilizado sin conocer sus efectos específicos en cada situación concreta de un contenido en realidad virtual, contribuye a desorientar al espectador, lo cual refuta nuestra hipótesis acerca de su utilidad puesto que no ha asistido positivamente la comunicación audiovisual.

Estos resultados deben considerarse dentro de un marco situacional específico, remitido a las condiciones y características del experimento, el uso de ciertos estímulos, la utilización de un entorno técnico determinado, y la participación de un grupo de sujetos.

Mis contribuciones en este proyecto incluyen:

- El estudio y la planificación previos
- El desarrollo del entorno sonoro de los experimentos: el panning orientativo y el J-cut. La implementación de los cambios necesarios en la adaptación del entorno utilizado por Serrano et al [26] en su investigación
- La preparación y realización de los experimentos
- La recopilación de datos y su posterior procesado, utilizando un código ya existente.
- El análisis de los datos y la elaboración de conclusiones

Personalmente, este trabajo me ha servido para aprender y ganar experiencia en diversas áreas vinculadas con el proyecto: cinematografía, realidad virtual, informática -especialmente programación en C#, manejo de programas de creación de contenido audiovisual como Unity3D y Adobe Premiere CC2017, toma y análisis de datos y su estadística, realización de pruebas de usuario, organización de proyectos... Me ha permitido utilizar una gran variedad de herramientas, estar en contacto con distintos campos de trabajo y me ha dado una visión amplia y veraz de lo que supone trabajar en un grupo de investigación. A lo largo del proyecto han aparecido problemas y obstáculos, pero han sido solventados y me he encontrado cómodo y a gusto durante todo el proceso, por lo que guardo una buena impresión y estoy dispuesto a seguir un camino similar en el futuro.

### 7.1 Implicaciones para el futuro

El conocimiento adquirido derivado de este trabajo, sumado al aportado por los autores citados y muchos otros que investigan actualmente en este campo, contribuye en su medida a asentar las bases de la edición cinematográfica en realidad virtual, a partir de las cuales se puede no sólo hacer cine, sino explorar nuevas alternativas e investigar en la misma u otras direcciones. De hecho, apenas se tiene información sobre este nicho y queda muchísimo camino que recorrer. Como dice Sebastian Sylwan [24]:

*“Como humanos tenemos el instinto de buscar nuevas formas de comunicación. Lo que hacemos es adaptar nuestro entorno, la tecnología que tenemos disponible, para conseguir nuevas experiencias. Y al final cada medio encuentra su nicho. Por eso estamos interesados en encontrar qué es lo que funciona bien en realidad virtual y qué no. Hay cosas para las cuales el cine es lo mejor y otras para las que un libro es insustituible.”*

Numerosos retos y caminos se pueden tomar con el objetivo de ayudar a construir el marco de referencia de la edición en RV. Para empezar, han de validarse otras técnicas de entre las utilizadas anteriormente en el cine tradicional, y se ha de continuar investigando acerca de la cognición humana y del impacto en ésta de la realidad virtual. Pero más concretamente, en nuestra línea de investigación, cabe preguntarse por las posibilidades que ofrece el *panning* orientativo en otros escenarios de esta misma plataforma –la RV-, siendo un caso particular la selección activa del contenido sonoro mediante la orientación voluntaria (o involuntaria) de la cabeza. En otras palabras, elegir lo que se quiere oír en función de a dónde se mira, con útiles aplicaciones como, por ejemplo, oír más claramente al personaje que se está mirando, estando inmerso virtualmente en una escena que consiste en una conversación entre varios personajes.

*“La utopía está en el horizonte, sé que nunca la alcanzaré. Sé que si yo camino diez pasos, ella se alejará diez pasos, porque ella se va alejando a medida que yo me acerco. Entonces, ¿para qué sirve la utopía? Para eso sirve, para caminar”*

*-Fernando Birri*

# Bibliografía

---

- [1] Espacio VR (2016). Fundación Telefónica. Madrid, España. Recuperado de <https://espacio.fundaciontelefonica.com/evento/espacio-realidad-virtual-vr/>
- [2] Innovation Enterprise (2017). Argyle Executive Forum. London, England. Recuperado de <https://channels.theinnovationenterprise.com/articles/is-vr-the-future-of-design>
- [3] Oculus (2017). Oculus VR. Spain. Recuperado de: <https://www.oculus.com>
- [4] Bordwell, D., Thompson, K., & Smith, J. (1997). Film art: An introduction (Vol. 7). New York: McGraw-Hill.
- [5] Hitchcock, A. (1948). Rope [La soga]. EEUU: Warner Bros. Pictures.
- [6] Iñárritu, A. G. (2914). Birdman (or the unexpected virtue of ignorance). EEUU: 20th Century Fox.
- [7] Cuarón, A. (2006). Children of men. RU, EEUU, Japan: Universal Picture, Toho-Towa, Strike Entertainment, Ingenious Film Partners, Hit & Run Productions.
- [8] Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, 6(1), 1-27.
- [9] Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), 355.
- [10] Bordwell, D., Staiger, J., & Thompson, K. (1985). *The classical Hollywood cinema: Film style & mode of production to 1960*. Columbia University Press.
- [11] Cutting, J. E. (2005). Perceiving scenes in film and in the world. *Moving image theory: Ecological considerations*, 9-27.
- [12] Gibson, E., & Rader, N. (1979). Attention. In *Attention and cognitive development* (pp. 1-21). Springer US.
- [13] Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5(4), 644-649.

- [14] Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9), 1059-1074. Recuperado de: [https://www.youtube.com/watch?time\\_continue=1&v=vJG698U2Mvo](https://www.youtube.com/watch?time_continue=1&v=vJG698U2Mvo)
- [15] Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80-84.
- [16] Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in human neuroscience*, 4.
- [17] Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489-1517.
- [18] Smith, T. J. (2006). An attentional theory of continuity editing. PhD thesis, University of Edimburg
- [19] Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin*, 133(2), 273
- [20] O'Steen, B. (2009). *The Invisible Cut*. Michael Wiese Productions.
- [21] Katz, S. D. (1991). Film directing shot by shot: visualizing from concept to screen. Gulf Professional Publishing.
- [22] Kon, S. (2006). *Papurika* [Paprika]. Japan, Tokyo: Sony.
- [23] Martínez-Salanova, E. El cine antes del cine: los antecedentes del cine. Recuperado de: <https://www.uhu.es/cine.educacion/cineyeducacion/cineprecine.htm>
- [24] Sebastian Sylwan, ABC Cultural, 22/08/2017. Recuperado de: [http://www.abc.es/cultura/abci-sebastian-sylwan-realidad-virtual-esta-como-cine-principios-siglo-pasado-201708220248\\_noticia.html](http://www.abc.es/cultura/abci-sebastian-sylwan-realidad-virtual-esta-como-cine-principios-siglo-pasado-201708220248_noticia.html)
- [25] Brillhart, J., Medium (2016). Recuperado de: <https://medium.com/@brillhart>
- [26] Serrano, A., Sitzmann, V., Ruiz-Borau, J., Wetzstein, G., Gutierrez, D., & Masia, B. (2017). Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (TOG)*, 36(4), 47.
- [27] Cameron, C., Vimeo. Recuperado de: <https://vimeo.com/blog/post/j-cuts-l-cuts>
- [28] Waltz for One, Intellectual Propaganda (2012), Vimeo. Recuperado de: <https://vimeo.com/44025740>
- [29] Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception*, 30(8), 993-1006.
- [30] Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews. Neuroscience*, 5(3), 229.
- [31] Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, 11(5), 5-5.

- [32] Wikimedia Commons, (2017). Wikipedia. Recuperado de.  
[https://es.wikipedia.org/wiki/Ojo#/media/File:Schematic\\_diagram\\_of\\_the\\_human\\_eye-es.svg](https://es.wikipedia.org/wiki/Ojo#/media/File:Schematic_diagram_of_the_human_eye-es.svg)
- [33] Pupil Labs, (2017). Recuperado de: <https://pupil-labs.com/vr-ar/>
- [34] Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.

## Anexos

---

## A. Recopilación de vídeos 360 o RV: enlaces

---

Aquí se muestra una recopilación de vídeos, películas o documentales para ser vistos en realidad virtual, que han sido utilizados durante la fase de investigación y de diseño de los experimentos para encontrar las claves relevantes en el estudio de la atención de los sujetos:

- HoneyVR (A few shorts, they look promising but still not released)  
<http://www.honeyvr.com/vr.films/>
- The invisible man – youtube360 (one shot filmed video)  
[https://www.youtube.com/watch?v=I\\_FUpUi2LBk](https://www.youtube.com/watch?v=I_FUpUi2LBk)
- The RooM – youtube360 (horror animated short)  
<https://www.youtube.com/watch?v=9vgBDiDpLmU>  
<https://www.youtube.com/watch?v=8JMkKC6o-To>
- Welcome – youtube360 <https://www.youtube.com/watch?v=REAwGmv0Fuk>
- LeMans – youtube360 <https://www.youtube.com/watch?v=LD4XfM2TZ2k>
- Earth Encounter – youtube360 <https://www.youtube.com/watch?v=mNVPckKOrdo>
- Time Couch – youtube360 <https://www.youtube.com/watch?v=7IaYJZ2Usdk>
- Bambino – youtube360 <https://www.youtube.com/watch?v=ujYyE0lfSUk>
- Night Light – youtube360 <https://www.youtube.com/watch?v=xfEnP2HGgRY>
- The fight for falluja – youtube360 [https://www.youtube.com/watch?v=\\_Ar0UkmID6s](https://www.youtube.com/watch?v=_Ar0UkmID6s)
- Mr. Robot 360 experience – youtube360  
<https://www.youtube.com/watch?v=H2Jc1wHlhEU>
- Mithbusters: Shark Shipwreck (Discovery VR) - youtube360  
<https://www.youtube.com/watch?v=aQd41nbQM-U>
- Star Wars: hunting of the Fallen - youtube360

<https://www.youtube.com/watch?v=SeDOoLwQQGo>

- Beyond the Night - youtube360 <https://www.youtube.com/watch?v=ScpwiEU9zI8>
- Shamans of Siberia - youtube360 <https://www.youtube.com/watch?v=rInUt3rJ3AE>

## B. Clips de video 360 representativos

---

Para la comparación del comportamiento de los sujetos de nuestro experimento con respecto a los participantes del experimento de Serrano et al. [26], además de generar las métricas se han generado una serie de vídeos que muestran (en proyección equirectangular) la trayectoria visual seguida por los usuarios proyectada sobre la imagen. Los clips están en baja resolución (los vídeos originales están en resolución de 3840x1920 px). En los obtenidos en este trabajo se puede apreciar la mayor afinidad de los sujetos, y la mayor rapidez en encontrar las ROIs.

Para acceder a ellos se debe usar el siguiente enlace: <https://goo.gl/3uJGai>

## C. Procesado de datos

---

### C.1 Recopilación de datos

Durante el experimento, Unity registra la orientación de la cabeza dada por el casco Oculus DK2, y la localización de la mirada tomada por los eye trackers de PupilCapture. Lo hace en una estructura de carpetas de forma que para cada sujeto, guarda esta información en dos archivos de Excel (uno para cada tipo de dato) en una carpeta con el nombre del clip visualizado. De esta forma podemos saber en qué punto de la escena de un vídeo particular estaba mirando cada sujeto (de forma anónima), y con qué precisión y confianza conocemos ese dato.

### C.2 Pre-procesado: baseline

Cuando durante un experimento se muestra un estímulo a un conjunto de personas, puede existir una mayor o menor variabilidad en el comportamiento de los sujetos ante dicho estímulo. Si se estudia el efecto de un determinado fenómeno durante la prueba, se necesitará conocer el comportamiento “base” de los sujetos a lo largo de la prueba, esto es, sin ese fenómeno concreto, para después poder compararlos ambos y apreciar las diferencias, si las hay.

Este comportamiento base, o *baseline*, se había obtenido previamente a este trabajo por Serrano et al [25] sometiendo a sujetos distintos de los de los experimentos a la visualización de los vídeos originales sin editar; es decir, sin cortes. Se había recopilado la información procediendo de la misma manera, mantenido las condiciones del experimento para asegurar su reproducibilidad. Al actuar de esta forma se elimina el efecto que la técnica de edición haya podido tener sobre el espectador, permitiendo comparar luego ambos comportamientos y apreciar la influencia del corte en la alteración de la atención.

Para realizar el baseline se visualizaron por orden aleatorio la totalidad de los vídeos originales por 10 personas diferentes. A partir del seguimiento ocular de las pupilas de los

sujetos se construyó una *trayectoria de mirada media* para cada vídeo –ver Figura C.1–, la cual, para corroborar su utilidad, debía cumplir que los datos recogidos fueran congruentes entre todos los observadores. Para comprobarlo se utilizó la herramienta llamada *curva ROC* (*Receiver Operating Characteristic*), que asigna, para cada instante, una medida de la congruencia entre distintos espectadores.

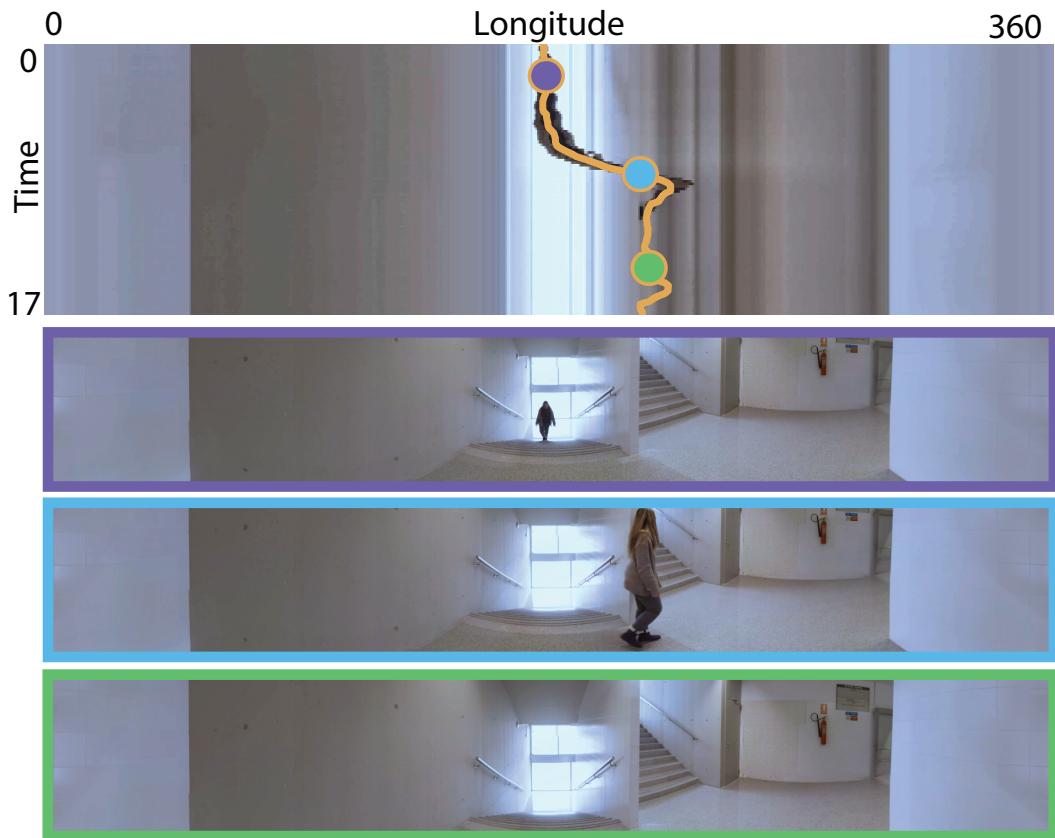


Figura C.1 Representación del baseline de una de las escenas de los vídeos 360

Esta curva funciona de la siguiente manera: inicialmente se construye un mapa de saliencia del vídeo en tramos de 2 segundos a partir de las fijaciones del total de los  $n$  usuarios (en este caso  $n=10$ ). Este mapa se divide en regiones con distinto percentil de saliencia, como las áreas encerradas por una curva de nivel de un mapa topográfico. Dejando un sujeto  $i$  fuera, con el resto de sujetos  $n-1$  se calcula, para el mapa obtenido de cada tramo de 2 segundos, las zonas con una saliencia determinada (percentil  $k$ ); después, tomando el sujeto  $i$ , se cuenta el número de fijaciones que ha realizado en esa zona. Este cálculo se realiza de nuevo para percentiles de valor  $k = \{0\%, \dots, 100\%\}$ , y el conjunto de los puntos determinados por el número de fijaciones en cada percentil conforma la curva.

Este proceso ha de realizarse con la totalidad  $n$  de sujetos, cada uno de los cuales da lugar a una curva distinta. Finalmente se realiza la media de todas ellas, dando como resultado la curva ROC; cuanto mayor sea el área contenida bajo la misma, mayor será la congruencia

### C. Procesado de datos

---

entre los sujetos, pues significará que sus fijaciones han caído en mayor porcentaje dentro de los percentiles de saliencia.

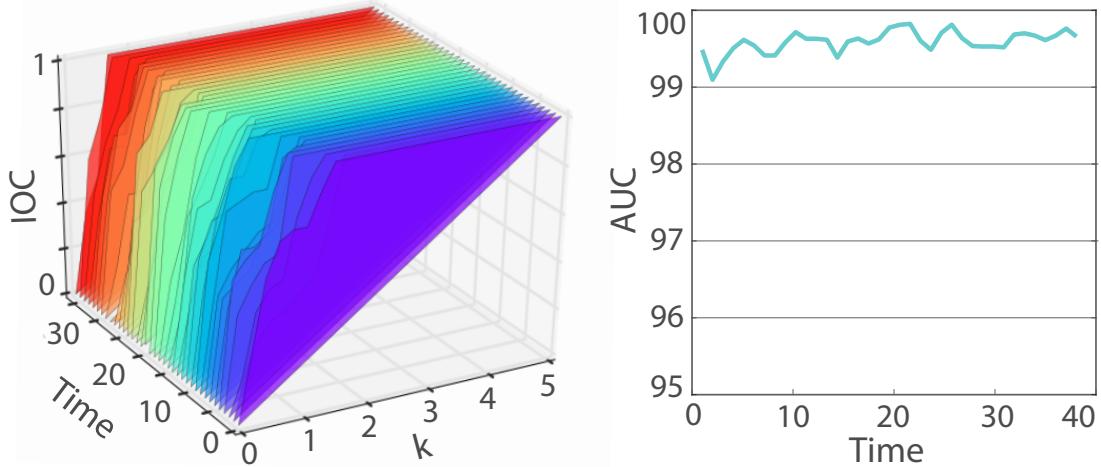


Figura C.2 Representación gráfica de la evolución temporal de la curva ROC

Dado que cada 2 segundos se ha calculado una curva, se puede representar temporalmente su evolución, que se muestra en la Figura C.2. Nos indica que se llega con muy pocos percentiles (2%) a la totalidad de la congruencia, y como es de esperar no desciende conforme estos aumentan. Por tanto, la baseline es válida.

### C.3 Tratamiento de los datos

Para preprocesar los datos se hace uso del código públicamente disponible del trabajo de Serrano et al.

Mediante un script programado en Python 2.7 se toman los archivos de Excel, se reordenan y se guardan en formato .mat, para que puedan ser leídos por *Matlab R2015*, una herramienta de software matemático que aporta un IDE (entorno de desarrollo integrado) con su propio lenguaje. Entre las funciones de este programa se encuentran la interacción con otros programas en sus respectivos lenguajes y con otros elementos de hardware, la representación gráfica de datos y funciones, la compilación de algoritmos y la creación de GUIs (interfaces gráficas de usuario).

Mediante una serie de scripts en Matlab se recogen todos los tests realizados y se asigna una muestra de la mirada tomada por el eyetracker a cada fotograma del vídeo. Genera de nuevo los clips de vídeo, pero con los puntos de la imagen a los que se ha mirado durante el experimento pintados de colores, con un código de color por sujeto. De esta forma podemos visualizar el recorrido de la mirada de los observadores sobre los fotogramas.

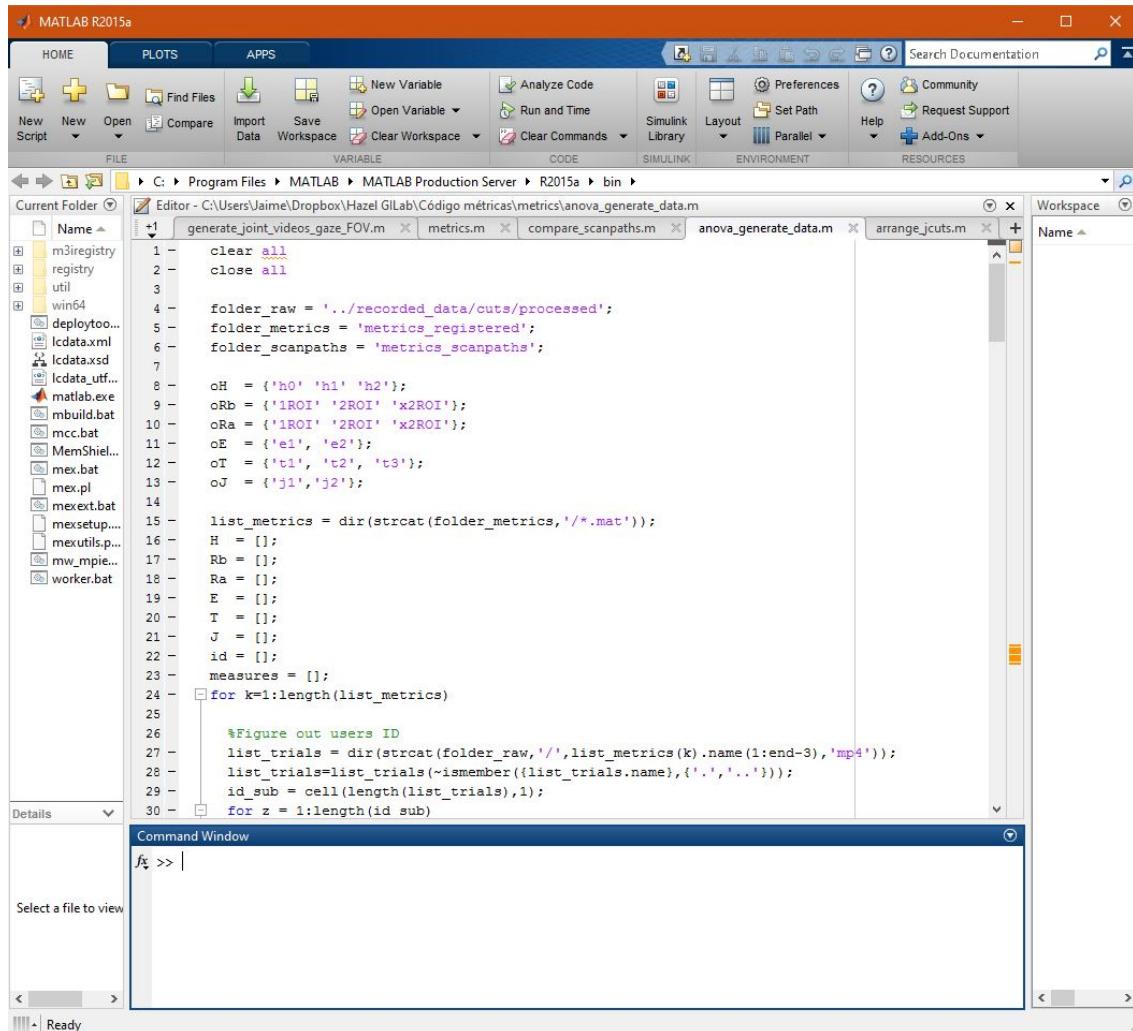


Figura X.X Interfaz de Matlab

Una vez colocados los archivos según la estructura que dicta el script de Matlab, se ejecuta éste y calcula las métricas deseadas, clasificándolas en un Excel de forma que sean cómodamente utilizables en *SPSS17*, un software dedicado al análisis estadístico y tratamiento de datos. Con este programa se obtienen los valores medios de las medidas, su desviación y se someten a un análisis de efectos combinados (*Mixed Model Analysis*) para determinar qué métricas se ven afectadas por las variables, y de qué forma. Si el valor de significancia (Sig.) de las variables sobre cada parámetro en los Type III tests of Fixed Effects es inferior a 0,05, se considera que esa variable supone una diferencia significativa en el parámetro, y se pasa a estudiar el valor medio.

Para aprender a usar SPSS se utilizó el libro “*Discover Statistics Using SPSS*”, de Andy Field, 3<sup>a</sup> edición.

## C. Procesado de datos

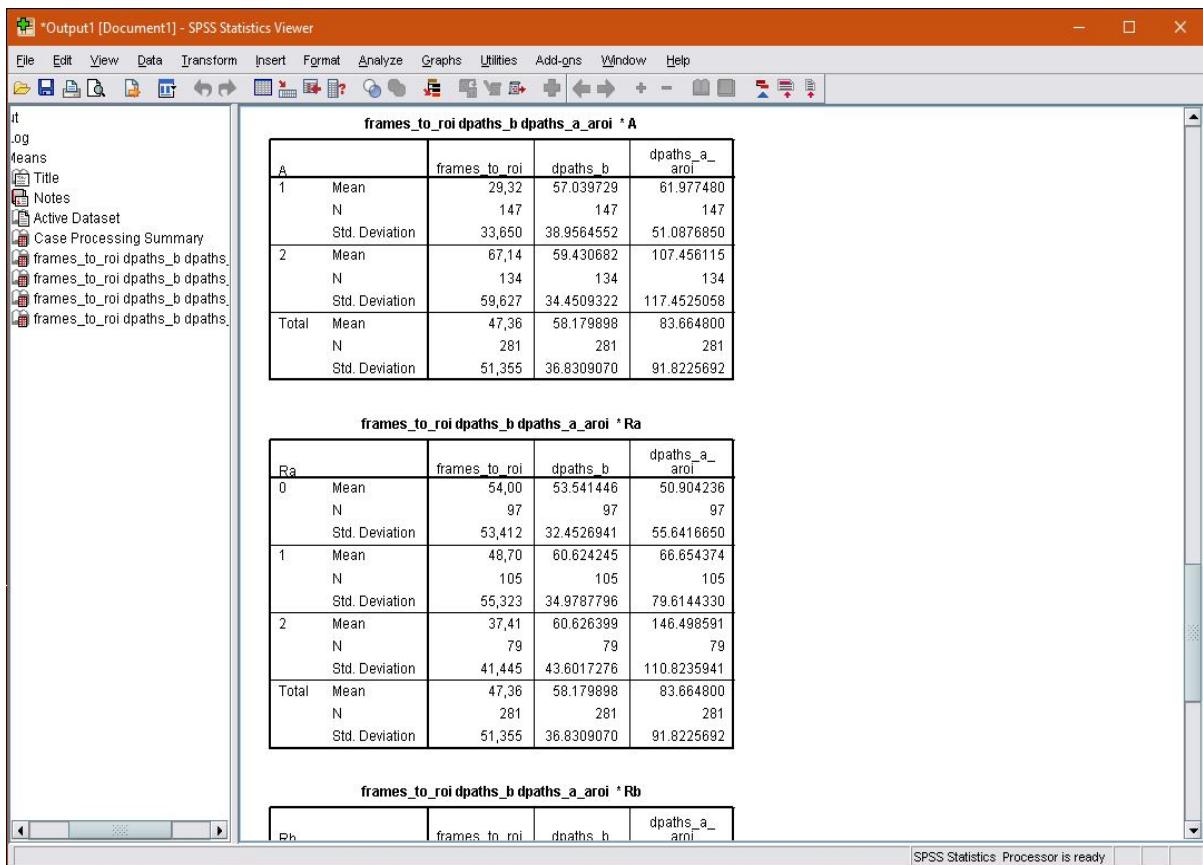


Figura 5.X Interfaz gráfica de SPSS

Para exemplificar el tipo y la forma de la información que devuelve SPSS, a continuación se añaden algunas de las tablas de resultados. En concreto, se muestra el *Mixed Model Analysis* para la métrica *Fotogramas hasta encontrar una ROI*.

Es interesante comprobar el valor de significancia (Sig.) de las variables sobre el parámetro *Fotogramas hasta encontrar una ROI* en los Type III tests of Fixed Effects. Por ejemplo, podemos ver que efectivamente, como se ha dicho en el Capítulo 6. Resultados, la variable A (A80) demuestra ser significativa, y por ello se estudia y compara su media.

```

DATASET ACTIVATE DataSet4.

* Encoding: UTF-8.

MIXED framesToROI WITH Dir A80 Ra1 Ra2 Rb1 Rb2
  /CRITERIA=CIN(95) MXITER(800) MXSTEP(10) SCORING(1) SINGULAR(0.0000000000
01) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=
    A80 Ra1 Ra2 Rb1 Rb2 Dir
    A80*Ra1 A80*Ra2 A80*Rb1 A80*Rb2 A80*Dir
    Ra1*Rb1 Ra1*Rb2 Ra1*Dir
    Ra2*Rb1 Ra2*Rb2 Ra2*Dir
    Rb1*Dir Rb2*Dir
  | SSTYPE(3)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(VC) .

```

## Mixed Model Analysis

### Notes

Output Created		21-SEP-2017 15:46:36
Comments		
Input	Data	D: \Users\Ana\Dropbox\RESEARCH\PFCs_dirigidos\Javier Camon\exp1_new-sigg.sav
	Active Dataset	DataSet4
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	887
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data for all variables in the model.

### Information Criteria<sup>a</sup>

-2 Log Likelihood	8819,964
Akaike's Information Criterion (AIC)	8863,964
Hurvich and Tsai's Criterion (AICC)	8865,136
Bozdogan's Criterion (CAIC)	8991,297
Schwarz's Bayesian Criterion (BIC)	8969,297

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: framesToROI.

### Fixed Effects

#### Type III Tests of Fixed Effects<sup>a</sup>

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	753,283	41,579	,000
A80	1	880,000	116,071	,000
Ra1	1	878,531	1,114	,292
Ra2	1	883,620	,162	,687
Rb1	1	885,758	,299	,584
Rb2	1	877,485	2,704	,100
Dir	1	494,027	8,402	,004
A80 * Ra1	1	884,222	2,626	,105
A80 * Ra2	1	879,056	8,680	,003
A80 * Rb1	1	886,877	6,863	,009
A80 * Rb2	1	884,602	17,027	,000
Dir * A80	1	876,443	6,910	,009
Ra1 * Rb1	1	866,739	,922	,337
Ra1 * Rb2	1	865,374	2,827	,093
Dir * Ra1	1	865,212	,694	,405
Ra2 * Rb1	1	876,236	1,018	,313
Ra2 * Rb2	1	871,224	3,356	,067
Dir * Ra2	1	864,854	5,559	,019
Dir * Rb1	1	874,723	3,428	,064
Dir * Rb2	1	864,255	,059	,809

a. Dependent Variable: framesToROI.

### Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% ...
						Lower Bound
Intercept	31,750950	4,923993	753,283	6,448	,000	22,084569
A80	59,746457	5,545621	880,000	10,774	,000	48,862270
Ra1	-6,480881	6,140629	878,531	-1,055	,292	-18,532897
Ra2	-2,447196	6,074991	883,620	-,403	,687	-14,370292
Rb1	3,395062	6,205615	885,758	,547	,584	-8,784361
Rb2	10,033626	6,102091	877,485	1,644	,100	-1,942772
Dir	-17,704983	6,107911	494,027	-2,899	,004	-29,705670
A80 * Ra1	-9,417579	5,811360	884,222	-1,621	,105	-20,823247
A80 * Ra2	-16,765730	5,690720	879,056	-2,946	,003	-27,934715
A80 * Rb1	-15,142265	5,780078	886,877	-2,620	,009	-26,486491
A80 * Rb2	-23,572318	5,712582	884,602	-4,126	,000	-34,784114
Dir * A80	-12,841794	4,885204	876,443	-2,629	,009	-22,429859
Ra1 * Rb1	-6,824860	7,106826	866,739	-,960	,337	-20,773461
Ra1 * Rb2	11,852283	7,049237	865,374	1,681	,093	-1,983319
Dir * Ra1	5,020085	6,027834	865,212	,833	,405	-6,810802
Ra2 * Rb1	-7,055779	6,991859	876,236	-1,009	,313	-20,778525
Ra2 * Rb2	-12,674181	6,917989	871,224	-1,832	,067	-26,252053
Dir * Ra2	13,950904	5,916867	864,854	2,358	,019	2,337805
Dir * Rb1	11,080556	5,984305	874,723	1,852	,064	-,664718
Dir * Rb2	-1,434710	5,929182	864,255	-,242	,809	-13,071990

### Estimates of Fixed Effects<sup>a</sup>

Parameter	95% Confidence .
	Upper Bound
Intercept	41,417331
A80	70,630644
Ra1	5,571135
Ra2	9,475900
Rb1	15,574486
Rb2	22,010024
Dir	-5,704296
A80 * Ra1	1,988089
A80 * Ra2	-5,596746
A80 * Rb1	-3,798039
A80 * Rb2	-12,360522
Dir * A80	-3,253730
Ra1 * Rb1	7,123741
Ra1 * Rb2	25,687884
Dir * Ra1	16,850972
Ra2 * Rb1	6,666968
Ra2 * Rb2	,903691
Dir * Ra2	25,564002
Dir * Rb1	22,825829
Dir * Rb2	10,202570

a. Dependent Variable: framesToROI.

### Covariance Parameters

#### Estimates of Covariance Parameters<sup>a</sup>

Parameter	Estimate	Std. Error	Wald Z	Sig.
Residual	1203,821006	58,700749	20,508	,000
Intercept [subject = ID] Variance	17,163695	17,176359	,999	,318

#### Estimates of Covariance Parameters<sup>a</sup>

Parameter	95% Confidence Interval	
	Lower Bound	Upper Bound
Residual	1094,096445	1324,549604
Intercept [subject = ID] Variance	2,414244	122,022635

a. Dependent Variable: framesToROI.

## D. Encuesta realizada en el experimento

---

EXPERIMENT ID \_\_\_\_\_

Demographics			
Age _____			
Gender _____			
Have you used a VR headset before?	YES / NO		
If YES: Do you use VR regularly (more than once a month)?	YES / NO		
Do you have eye strain, headaches and/or nausea in VR?	YES / NO		
Other information? _____			
Information about visual correction			
Have you had Lasik eye surgery?	YES / NO		
Do you have surgically implanted intraocular lenses (IOL)?	YES / NO		
Do you typically wear glasses or contacts?	YES / NO		
If YES:	For near vision? YES / NO		
	For far vision? YES / NO		
	What are you currently wearing? GLASSES / CONTACTS / NEITHER		
Other information? _____			
Presbyopia			
Presbyopia is the inability to focus on nearby objects associated with normal aging, often corrected with bifocals, progressive lenses, or monovision.			
Are you aware of having a diagnosis of presbyopia?	YES / NO		
If YES:	Do you wear monovision correction? YES / NO		
	If YES, which of your eyes you use for near vision? LEFT / RIGHT		
If not monovision, how is your presbyopia corrected? _____			
Other information? _____			
Ocular Health			
Are you aware of having a diagnosis of:			
Amblyopia	Strabismus/lazy eye	Glaucoma	Diabetic Retinopathy
Cataracts	Light sensitivity	Nystagmus	Macular Degeneration
Other information? _____			
After the test			
Did you experience visual discomfort (such as headache, eye strain...)?			
Did you experience dizziness or motion sickness?			
Did you see any artifacts in the videos?			
Did you understand the actions being performed?			
Did you at any point feel lost when trying to follow the actions?			
Were the videos comfortable to watch (e.g., they did not require too much head movement)?			

## E. Espacio de variables

---

Las combinaciones del desalineamiento A ( $A_{40}$  y  $A_{80}$ ) y la configuración de ROIs, Rb y Ra, dan como resultado el siguiente espacio de 18 posibilidades:

$A_{40}$	$R_{b,0}$	$R_{a,0}$	$(A_{40}, R_{b,0}, R_{a,0})$
		$R_{a,1}$	$(A_{40}, R_{b,0}, R_{a,1})$
		$R_{a,2}$	$(A_{40}, R_{b,0}, R_{a,2})$
	$R_{b,1}$	$R_{a,0}$	$(A_{40}, R_{b,1}, R_{a,0})$
		$R_{a,1}$	$(A_{40}, R_{b,1}, R_{a,1})$
		$R_{a,2}$	$(A_{40}, R_{b,1}, R_{a,2})$
	$R_{b,2}$	$R_{a,0}$	$(A_{40}, R_{b,2}, R_{a,0})$
		$R_{a,1}$	$(A_{40}, R_{b,2}, R_{a,1})$
		$R_{a,2}$	$(A_{40}, R_{b,2}, R_{a,2})$
$A_{80}$	$R_{b,0}$	$R_{a,0}$	$(A_{80}, R_{b,0}, R_{a,0})$
		$R_{a,1}$	$(A_{80}, R_{b,0}, R_{a,1})$
		$R_{a,2}$	$(A_{80}, R_{b,0}, R_{a,2})$
	$R_{b,1}$	$R_{a,0}$	$(A_{80}, R_{b,1}, R_{a,0})$
		$R_{a,1}$	$(A_{80}, R_{b,1}, R_{a,1})$
		$R_{a,2}$	$(A_{80}, R_{b,1}, R_{a,2})$
	$R_{b,2}$	$R_{a,0}$	$(A_{80}, R_{b,2}, R_{a,0})$
		$R_{a,1}$	$(A_{80}, R_{b,2}, R_{a,1})$
		$R_{a,2}$	$(A_{80}, R_{b,2}, R_{a,2})$

## F. Código desarrollado

---

Se presenta aquí un fragmento del código empleado en los scrpts de Unity3D, con el fin de exemplificar el trabajo realizado. Se ha elegido este código porque es representativo, al ser el encargado de reproducir el sonido con panning orientativo; está entegrado en el script controlador que maneja la escena.

En primer lugar se introducen las clases, luego algunas variables globales relacionadas con el fragmento de código a mostrar, y después se puede seguir **speaker\_setup**, la función encargada de la distribución de los speakers en la escena de Unity3D. Finalmente aparece una porción de código destinada a detectar el corte en los vídeos y llamar a speaker\_setup cuando se produce el corte cinematográfico, con el fin de que la función vuelva a colocar los speakers donde se encuentran los nuevos personajes.

```
using UnityEngine;
using System.Collections;
using UnityEditor;
using System.IO;
using System;
using System.Threading;
using UnityOSC;
using System.Collections.Generic;
using System.Data;
using UnityEngine.SceneManagement;
using RenderHeads.Media.AVProVideo;

public class CameraTracker : MonoBehaviour
{
    // Para elegir el adelanto del J-cut/For choosing the J-cut delay
    public float delayJcut = 0f;

    // momento del corte./Allows to change the moment of removement of
    // speakers in the cut, depending on the video experiment
    public float cut_time = 6f;

    // For random J-cut delay {0.5, 1, 2}
    public bool all_delays = false;
```

```

    // Aux variable to use the name of the current video as a bridge from
    cuts[i].key to "speaker_setup"
    private string nameVar;

    // Aux variable to use the time of the current video as a bridge from
    elapsed_time to "speaker_setup"
    private float timeVar = 0f;

    // Aux variable to keep the actual delay of the video to use it in the
    cut
    private float actualDelay = 0;

    // Aux variable to place only once the speakers in the cut
    private bool setup_done = false;

}

[...]

//ESTA FUNCION COLOCA LOS SPEAKERS EN EL ESPACIO DE UNITY EN FUNCIÓN
DEL VIDEO ESCOGIDO DE FORMA RANDOM
//THIS FUNCTION PLACES SPEAKER IN THE VIRTUAL SPACE OF UNITY DEPENDING
ON THE RANDOMLY CHOSEN VIDEO

private void speaker_setup(string videoName, float videoTime)
{
    GameObject speaker1 = GameObject.Find("Speaker1");
    GameObject speaker2 = GameObject.Find("Speaker2");

    if (0.5 > videoTime)
    {
        //Cogemos los speakers/We get the speakers 1 & 2 here
        AudioSource audio1 = speaker1.GetComponent<AudioSource>();
        AudioSource audio2 = speaker2.GetComponent<AudioSource>();

        //Le dice a los speakers qué clip tienen que cargar
        //Tells the speakers which clips they have to load
        string clipName1 = videoName.Substring(3, 13);
        string clipName2 = videoName.Substring(3, 13);

        actualDelay = delayJcut;

        if (delayJcut == 0)
        {
            clipName1 += "_s1";
            clipName2 += "_s2";
        }

        else if (delayJcut == 1)
        {
            clipName1 += "_s1_j1";
            clipName2 += "_s2_j1";
            if (all_delays)
            {
                delayJcut = 2;
            }
        }

        else if (delayJcut == 2)
    }
}

```

```

    {
        clipName1 += "_s1_j2";
        clipName2 += "_s2_j2";
        if (all_delays)
        {
            delayJcut = 1;
        }
    }

    print("El delay es " + actualDelay); //debug

    audio1.clip = Resources.Load<AudioClip>("ImportedAudios/" +
clipName1);
    audio2.clip = Resources.Load<AudioClip>("ImportedAudios/" +
clipName2);

        // Reproduce los audios desde la fuente/Plays both audio clips
from Audio Source

GameObject.Find("Speaker1").GetComponent< AudioSource >().PlayDelayed(0);

GameObject.Find("Speaker2").GetComponent< AudioSource >().PlayDelayed(0);
}

//Esto lee la lista de coordenadas
//All this below reads a cliplist.txt to input the coordinates of
the speakers:
//Directamente del archivo de texto/Read the text directly from the
cliplist.txt file
string path = "Assets/Resources/cliplist.txt";
StreamReader reader = new StreamReader(path);

/*This piece of code opens up a file, reads its lines until it
finds the one with the video name,
and uses the Split method with the correct delimiter to store the
line in the parts[].
StringSplitOptions is used to tell the Split method to ignore empty
spaces within the splits.*/
string[] parts = new string[13];
while ((path = reader.ReadLine()) != null)
{
    char[] delimiter = new char[] { '\t' };
    parts = path.Split(delimiter,
StringSplitOptions.RemoveEmptyEntries);

    if (parts[0] == videoName)
    {
        break;
    }
}
reader.Close();

// Coloca los speakers dependiendo del momento
//Place the speakers depending on the moment: before or after the
cut
if ((cut_time-actualDelay) > videoTime)
{ // From string to float

```

```

        float[] coord = new float[6];
        for (int i = 0; i < 6; i++)
        {
            coord[i] = float.Parse(parts[i+1]);
        }

        // INITIAL SPEAKERS SETUP
        Vector3 position1 = new Vector3(coord[0], coord[1], coord[2]);
        Vector3 position2 = new Vector3(coord[3], coord[4], coord[5]);

        speaker1.transform.position = position1;
        speaker2.transform.position = position2;

    }

    else
    { // From string to float
        float[] coord = new float[6];
        for (int i = 0; i < 6; i++)
        {
            coord[i] = float.Parse(parts[i+7]);
        }

        // AFTER CUT SPEAKERS SETUP
        Vector3 position1 = new Vector3(coord[0], coord[1], coord[2]);
        Vector3 position2 = new Vector3(coord[3], coord[4], coord[5]);

        speaker1.transform.position = position1;
        speaker2.transform.position = position2;

        setup_done = true;
    }
}

[...]

// Mientras el video continue/While the video hasn't ended, do
this:
while (duration > elapsed_time)
{
    Log(elapsed_time);
    yield return null; //Estos yields paran corrutinas

    //En el momento del corte llama a speaker_setup para actualizar
el speaker
    //If the time is the moment of cut, it calls "speaker_setup()"
for cut speaker new replacement
    if ((cut_time-actualDelay) < elapsed_time && !setup_done)
    {
        timeVar = elapsed_time;
        speaker_setup(nameVar, timeVar);
    }

    elapsed_time += Time.deltaTime; //Acumula el tiempo/Sums the
elapsed time to count the time in seconds
}

```





