



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ANÁLISIS DE SEGUIMIENTO OCULAR DE GRUPOS ETARIOS CHILENOS EN
PÁGINAS WEB DE ECOMMERCE PARA EVALUAR LA GENERACIÓN DE UN
MODELO PREDICTIVO ALTERNATIVO DIFERENCIADOR AL PROYECTO AKORI

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

BASTIÁN AARON PARDO VERGARA

PROFESOR GUÍA:
JUAN VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:
IGNACIO CALISTO LEIVA
FELIPE VERA CID

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: BASTIÁN AARON PARDO VERGARA
FECHA: 2019
PROF. GUÍA: JUAN VELÁSQUEZ SILVA

ANÁLISIS DE SEGUIMIENTO OCULAR DE GRUPOS ETARIOS CHILENOS EN PÁGINAS WEB DE ECOMMERCE PARA EVALUAR LA GENERACIÓN DE UN MODELO PREDICTIVO ALTERNATIVO DIFERENCIADOR AL PROYECTO AKORI

Del constante crecimiento de las compras a través de la web en Chile surge la necesidad en las empresas ofertantes de transformar sus canales de venta, dando prioridad a la compra en línea. Resultado de esto, es fundamental que sus sitios web estén optimizados adecuadamente representando de mejor manera lo que ellos quieren transmitir y dando énfasis a lo que de verdad es importante para ellos.

Este trabajo de título tiene como objetivo general comparar patrones de visualización de páginas web de eCommerce entre grupos etarios chilenos mediante análisis estadístico para evaluar la generación de un modelo clasificador de rendimiento superior al del proyecto Akori. Para lo anterior, se hace un experimento con una herramienta de *Eye Tracking* recientemente adquirida por el Web Intelligence Centre para obtener datos reales de seguimiento ocular.

El proyecto Akori es un proyecto elaborado por el Web Intelligence Centre y el Laboratorio de Neuro Sistemas de la Universidad de Chile, que está desarrollando una plataforma informática para mejorar la estructura y contenido de sitios web mediante el uso de Web Intelligence y variables fisiológicas. En un futuro no lejano, Akori pretende comercializar su servicio a empresas de distintos rubros, entre ellos, eCommerce.

Dentro este proyecto, se han elaborado alrededor de 7 memorias de investigación de estudiantes de pregrado del Departamento de Ingeniería Industrial de la Universidad de Chile, de las cuales se han podido validar o rechazar distintas hipótesis que relacionan la dilatación pupilar y la actividad bioeléctrica cerebral con el entorno web y que han abierto nuevas líneas de investigación. Para esta nueva etapa del proyecto Akori, se quiere estudiar el comportamiento visual de personas, utilizando sitios web reales. Es en este contexto que se desarrolla esta memoria cuyo fin es validar o rechazar la hipótesis de investigación que señala que “Es posible diferenciar significativamente las saliencias de páginas web de eCommerce para diferentes grupos etarios chilenos utilizando su comportamiento visual”. Una saliencia es una cualidad estética de la forma de un objeto o una configuración que destaca uno o más de sus rasgos por sobre los demás y que, en consecuencia, llama la atención.

Con los datos de seguimiento ocular de 80 personas, se cuantificaron tres patrones en las saliencias; la exploratividad, medida por la entropía, el acuerdo en las zonas observadas, medido por el *AUC-Judd* y el sesgo central, medido por la distancia euclideana entre el centro del mapa de saliencia y el de la imagen. Los resultados obtenidos son que no es posible hallar diferencias significativas en estas métricas para cada segmento de edad escogido por lo que no se puede generar un modelo clasificador a partir de estos patrones en la data obtenida del experimento. Por otro lado, se valida la precisión del modelo SAM-Resnet para su uso en pronóstico de saliencias en páginas web de eCommerce.

Para todas esas personas que están dispuestas a ayudar al otro sin esperar nada a cambio

Agradecimientos

A mi familia, a todos los participantes del experimento, al profesor Juan Velásquez por darme la oportunidad, a cada uno de los integrantes del Web Intelligence Centre y en particular a Felipe Vera por estar siempre ahí.

Tabla de Contenido

CAPITULO 1: INTRODUCCIÓN	1
1.1. Antecedentes Generales	1
1.1.1. Proyecto Akori	1
1.1.2. Identificación de la oportunidad	2
1.2. Preguntas de investigación e hipótesis	2
1.3. Descripción de la investigación	3
1.4. Objetivos	3
1.4.1. Objetivo general	3
1.4.2. Objetivos específicos	3
CAPITULO 2: ESTADO DEL ARTE	4
2.1. Trabajos de saliencias en el mundo	4
2.1.1. Inserción de publicidad basada en la atención	5
2.1.2. Marketing basado en la atención y optimización de la comunicación	6
2.1.2.1. Compañías basadas en algoritmos de saliencia	7
2.1.2.2. Sustitutos comerciales	8
CAPITULO 3: MARCO TEÓRICO	9
3.1. Pruebas de hipótesis	9
3.2. Estadística no paramétrica	12
3.2.1. Prueba de Levene	12
3.2.2. Prueba de Shapiro - Wilk	13
3.2.3. Prueba U Mann - Whitney	13
3.2.4. Prueba de Kruskal- Wallis	13
3.2.5. Prueba Posthoc de Dunn	14
3.3. Modelo de red neuronal SAM-Resnet	15
CAPITULO 4: METODOLOGÍA	16
4.1. Experimento	17
4.1.1. Instrumentación	17
4.1.2. Data	18
4.1.3. Metodología evaluación SAM-Resnet	18
4.2. Alcances	18
4.3. Selección imágenes experimento	19
4.3.1. Páginas eCommerce	19
4.4. Segmentos de edad	20

4.5.	Pre tratamiento de la data	20
4.6.	Medidas para cuantificar comportamiento visual	20
4.6.1.	Exploratividad	20
4.6.2.	Acuerdo	21
4.6.3.	Sesgo Central	22
4.6.4.	Correlación de Pearson (CC)	23
4.7.	Pruebas de hipótesis	23
CAPITULO 5: EXPERIMENTO Y RESULTADOS		25
5.1.	Procedimiento	25
5.1.1.	Selección de participantes	26
5.1.2.	Preprocesamiento datos seguimiento ocular	27
5.2.	Análisis de resultados	28
5.2.1.	Grupo experimental	29
5.2.2.	Correlación de Pearson	29
5.2.3.	Exploratividad	31
5.2.4.	Acuerdo	34
5.2.5.	Sesgo Central	39
5.2.6.	Mapas de Saliencia imágenes por segmento	40
5.3.	Evaluación modelo de saliencia SAM-Resnet	44
5.3.1.	Comparación gráfica de saliencias Ground Truth vs SAM-Resnet	46
CAPITULO 6: CONCLUSIONES		50
6.1.	Conclusión Investigación	50
6.2.	Potencial Comercial	52
6.3.	Recomendaciones y Trabajo Futuro	53
GLOSARIO		56
BIBLIOGRAFÍA		60
ANEXOS		69
8.1.	Datos demográficos participantes experimento	69
8.2.	Consentimiento Informado	70
8.3.	Páginas Web Experimento	71

Índice de Tablas

2.1. <i>Benchmarking</i> empresas sustitutas comerciales	8
4.1. Escenarios posibles en el transcurso del trabajo de título	16
4.2. Especificaciones de Hardware Tobii X2-30 [1]	17
5.1. Estadísticas descriptivas para la edad de los participantes del experimento .	29
5.2. Correlaciones de Pearson para cada combinación de segmentos en cada imagen	30
5.3. Estadísticas descriptivas CC Inter - Segmentos de edad y resultados de pruebas de hipótesis	30
5.4. Prueba de Levene conjuntos de entropía por segmento por imagen Enfoque 1	31
5.5. Prueba Kruskal conjuntos de entropía por segmento por imagen Enfoque 1 .	31
5.6. Matriz entropía Enfoque 2 y estadísticas descriptivas entropía	32
5.7. Pruebas estadísticas entropía	33
5.8. Entropía para mapas de saliencia promedio por segmento para cada imagen Enfoque 2	33
5.9. Estadísticas descriptivas Entropía para cada segmento Enfoque 2	34
5.10. <i>AUC</i> en donde el <i>source</i> es el mapa saliencia promedio de cada segmento y el <i>target</i> es el mapa de saliencia de cada sujeto para una imagen en particular .	35
5.11. Matriz resultante del promedio de las 18 matrices de <i>AUC</i>	35
5.12. Estadísticas descriptivas <i>AUC</i> por combinaciones de segmentos	35
5.13. Resultados pruebas estadísticas comparación <i>AUC</i> Intra - Segmentos de edad	36
5.14. Prueba de Levene y Kruskal Wallis <i>AUC</i> Enfoque 2 Intra grupos para cada imagen por separado	38
5.15. Prueba de Levene y Kruskal Wallis <i>AUC</i> Enfoque 2 Intra grupos	38
5.16. Pruebas estadísticas Sesgo Central	39
5.17. <i>AUC-Judd</i> para conjunto de imágenes 1	44
5.18. <i>AUC-Judd</i> para conjunto de imágenes 2	45
5.19. Estadísticas descriptivas <i>AUC</i> SAM-Resnet para el conjunto de todas las páginas web	45
5.20. Correlaciones de Pearson conjunto de imágenes 1	45
5.21. Correlaciones de Pearson conjunto de imágenes 2	45
5.22. Estadísticas descriptivas <i>CC</i> SAM-Resnet para el conjunto de todas las páginas web	46
8.1. Datos demográficos participantes experimento	69

Índice de Ilustraciones

4.1. Eye Tracker utilizado en el estudio	18
4.2. Valores de entropía de distintos sujetos para una misma imagen	21
4.3. Ejemplo curva ROC y cálculo de <i>AUC</i>	21
5.1. Pop-out utilizado en experimento	26
5.2. Mapa de saliencia vs Mapa de calor	28
5.3. Ejemplo mapa de saliencia promedio para una imagen particular	28
5.4. Gráfico de distribución de las edades de los participantes del experimento . .	29
5.5. Distribución Correlación de Pearson	30
5.6. Intervalos de confianza para entropía por segmentos en las imágenes que se encuentran diferencias significativas	32
5.7. Entropía Enfoque 1	33
5.8. Entropía para diferentes enfoques	34
5.9. Comparación gráfica <i>AUC</i> Intra e Inter segmentos	36
5.10. Resultados prueba de Dunn y distribución <i>AUC</i> Inter-Segmentos de edad .	37
5.11. Intervalos de confianza distancia euclidean por segmentos para imágenes 5, 6, 7, 9 y 12 respectivamente	39
5.12. Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente	41
5.13. Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente	42
5.14. Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente	43
5.15. Distribución de <i>AUC</i> modelo SAM-Resnet para 10 y 5 segundos de visualización considerando los dos conjuntos de imágenes	45
5.16. Distribución de <i>CC</i> modelo SAM-Resnet para 10 y 5 segundos de visualización considerando los dos conjuntos de imágenes	46
5.17. Por fila, Ground-Truth 10 y 5 segundos de visualización y saliencia entregada por SAM-Resnet respectivamente para una imagen en particular Conjunto 1	47
5.18. Por fila, Ground-Truth 10 y 5 segundos de visualización y saliencia entregada por SAM-Resnet respectivamente para una imagen en particular Conjunto 1	48
5.19. Por fila, dos pares de imágenes Ground Truth vs SAM-Resnet respectivamente, Conjunto 2 de imágenes	49
6.1. Ejemplo etiqueta en saliencia. Fuente: Adaptado de [2]	54
8.1. Consentimiento informado pag 1	70
8.2. Consentimiento informado pag 2	71
8.3. Conjunto de imágenes 1, web Casa & Ideas, Falabella, Jumbo, Dafiti, Despegar, Groupon	72

8.4. Conjunto de imágenes 2	73
8.5. Sala experimento	74

CAPITULO 1: INTRODUCCIÓN

1.1. Antecedentes Generales

1.1.1. Proyecto Akori

Akori es un proyecto que se ha desarrollado desde el año 2012 entre las facultades de Ciencias Físicas y Matemáticas y de Medicina de la Universidad de Chile.

Su nombre es el resultado de un acrónimo que significa *Advanced Kernel for Ocular Research and Web Intelligence* y además significa Halcón en mapudungún.

Este proyecto, se basa en un algoritmo de redes neuronales convolucionales, llamado *SAM-Resnet* desarrollado por Cornia et al [3], el cual al entregarle como *input* una imagen, mide las características de los distintos puntos de estas: color, contraste, luminosidad, caras, ejes, entre otras. Luego calcula, dado el conjunto de características de la imagen completa, cuáles son los puntos que más llaman la atención, es decir, aquellos con mayor saliencia visual, simulando la visión natural de las personas.

El enfoque del proyecto ha sido en investigación y desarrollo en tres áreas distintas: *Web Intelligence*, exploración visual y neurociencia. Durante los años de desarrollo del proyecto, hasta la actualidad, se han obtenido diversos resultados producto de la investigación continua en las áreas de estudio, decantando en varias publicaciones dentro de los distintas aristas de la investigación, entre los que se encuentran los siguientes tópicos: Predicción de intención de clic en usuarios [4] [5] [6], identificación de objetos claves dentro de una página web [7] [8], mejora en la identificación de objetos clave utilizando tecnologías de *Eye Tracking* [9] [10] y *EEG* [11], identificación de palabras clave dentro de una página web [12].

Actualmente, se cuenta con un prototipo funcional para páginas web, en el cual se carga una página durante un segundo, se obtiene una imagen de la página cargada y se realiza el análisis de la imagen. La atención de las personas sobre la imagen analizada es representada a través de un mapa de calor, que muestra los puntos de mayor atención en colores cálidos, mientras los que tienen menor probabilidad de llamar la atención aparecen en colores fríos. Otra forma de mostrarlo es a través de mapas de focos de atención, en donde se oscurece la imagen completa y los puntos o focos de atención aparecen claros. Un tercer mapa es el de saturación, en donde se muestran en claro las zonas más saturadas y en negro las que no lo están [13].

1.1.2. Identificación de la oportunidad

La situación actual del proyecto Akori es que ofrece un servicio a modo de prototipo que consta de un algoritmo de redes neuronales que recibe como *input* una imagen y entrega como output dicha imagen con un mapa de fijación ocular, mapa de foco y mapa de objetos claves. El problema presente hoy en día es que los mapas de fijación ocular que el algoritmo predice de una imagen (saliencias), no son posible diferenciarlos por grupo etario, sexo u otra variable demográfica, provocando que estos sean genéricos y carentes de información detallada.

Algunos de los objetivos a corto plazo del proyecto Akori son, por un lado, precisar de una mejor manera las saliencias que predice para tener una mejor exactitud y por otro, poder clasificarlas para así poder entregar algo más, agregando valor a su servicio y diferenciándose de la competencia.

Este vacío presente se pretende resolver mediante el reentrenamiento de la red neuronal con data de seguimiento ocular específica de usuarios chilenos y/o la propuesta de nuevos modelos predictivos o híbridos de los existentes hoy en día en la literatura.

Con el cumplimiento de estos objetivos Akori podría entregar un servicio más completo a sus clientes, siendo una herramienta útil para que ellos puedan alinear sus páginas web con su modelo de negocios y tomar acción en caso de que sus diseños no estén transmitiendo a sus usuarios lo que se desea transmitirles.

Recientemente el Web Intelligence Centre ganó un fondo FONDECYT del cual se destinaron recursos para la investigación en el proyecto Akori. Se adquirió un instrumento de *Eye tracker* modelo Tobii x2-30 [1] y un dispositivo galvánico con el fin de realizar experimentos con datos fisiológicos de personas y obtener revelaciones con estos para el mejoramiento del modelo *SAM-Resnet*.

1.2. Preguntas de investigación e hipótesis

A continuación, se enuncian las hipótesis que motivan el trabajo de memoria en cuestión.

1. “Es posible diferenciar significativamente las saliencias de páginas web de eCommerce para diferentes grupos etarios chilenos utilizando su comportamiento visual”
2. “El rendimiento del modelo *SAM - Resnet* en las páginas web de eCommerce seleccionadas para el experimento, medido por el *AUC-Judd*, está estadísticamente por debajo que el declarado en el *MIT Saliency Benchmark results: cat2000*” [14]

1.3. Descripción de la investigación

1.4. Objetivos

1.4.1. Objetivo general

“Comparar patrones de visualización de páginas web de eCommerce entre grupos etarios chilenos mediante análisis estadístico para evaluar la generación de un modelo clasificador de rendimiento superior al del proyecto Akori”

1.4.2. Objetivos específicos

1. Diseñar e implementar experimentos con el dispositivo Eye Tracker para la obtención de los datos a analizar
2. Evaluar rendimiento de modelo SAM-Resnet utilizado actualmente por el proyecto Akori en páginas webs de eCommerce
3. Validar o rechazar las hipótesis, en caso de ser validadas proponer un modelo para ser entrenado con la data obtenida y en caso que no, proponer una nueva línea investigativa para mejorar el modelo actual de Akori

El logro del objetivo general se mide a través de un experimento en donde se almacenan datos de seguimiento ocular de personas voluntarias y mediante pruebas estadísticas se obtiene el nivel de significancia de los resultados en cuanto a diferencias y similitudes en las saliencias de las páginas web escogidas.

CAPITULO 2: ESTADO DEL ARTE

2.1. Trabajos de saliencias en el mundo

Los factores que atraen la atención de un ser humano pueden estar relacionados con la escena observada o con el observador que la ve. Las características de abajo hacia arriba (*Bottom - up*) como el color, la intensidad y la orientación de una escena constituyen los factores relacionados con la escena, mientras que los factores relacionados con el observador se pueden clasificar como factores cognitivos y físicos (*Top - Down*).

La tendencia humana, el conocimiento previo, la emoción y la expectativa pueden asignarse como factores cognitivos, mientras que los factores físicos están predominantemente relacionados con el estado externo de los observadores humanos, como la edad del observador, la vista, la disparidad visual y el género.

La mayoría de los modelos de saliencia existentes se basan en factores relacionados con la escena [15] [16] [17] [18], es decir, las características propias de la imagen como color, tamaño, orientación, etc. En estos modelos, se obtiene un mapa de saliencia al integrar los mapas de características ascendentes en paralelo en diferentes escalas.

La teoría de integración de características de Treisman (FIT) [19] sirve de base para dichos modelos. Koulieris et al [20] analizaron en detalle los modelos de saliencia de bajo nivel y de alto nivel. Estudios relacionados con la visión [21] han revelado que el comportamiento de la mirada está estrechamente relacionado con los factores cognitivos de los observadores humanos, y algunos investigadores [22] [21] también intentaron modelar los factores cognitivos a través de los sistemas de atención computacional.

De manera similar, el papel de los factores físicos, especialmente los cambios en el desarrollo de la percepción visual y el control del movimiento ocular [23] [24] [25], ha sido estudiado extensivamente en psicología y neurobiología. Estos estudios han revelado diferencias relacionadas con la edad en cuanto a la conducta visual y el seguimiento visual. Krishna et al [26] han parametrizado las diferencias relacionadas con la edad en el comportamiento de visualización de la escena para cuatro grupos de edad diferentes (4 años, 6 años, 8 años y adultos de promedio 29 años) durante la visualización de escenas naturalistas tomadas de libros y películas infantiles. Informaron que la capacidad de exploración aumenta monótonamente con la edad (4 años hasta los adultos), mientras que el sesgo del centro es más alto entre los niños más pequeños (4 y 6 años). También propusieron un modelo adaptado por edad que se

puede usar para predecir las saliencias de niños y adultos.

Pruebas de apoyo de estudios de desarrollo [23] [24] sobre la exploración de la escena también ha demostrado que existen diferencias notables en el comportamiento de observación de la escena de los observadores en diferentes grupos de edad. Por ejemplo, se demostró que las características de la imagen local, como el color, la intensidad, la luminancia, etc., guían los aterrizajes de fijación en una etapa temprana de la vida, mientras que más adelante, los aterrizajes de fijación están dominados por un procesamiento más descendente (*Top-Down*) [23] [24].

En los últimos años, los estudios de investigación sobre la visión han estudiado el papel de los factores físicos del ser humano, como la edad, la disparidad visual, la vista y el género, para atraer la atención del ser humano durante la escena. Los estudios de desarrollo sugieren que los cambios relacionados con la edad en el control del movimiento ocular, como la capacidad para fijarse en el objetivo, mejoran ampliamente durante la primera infancia [27] [25] [28] [29] pero los aspectos más complejos del sistema de fijación, como la estabilidad de las fijaciones y el control cognitivo, continúan desarrollándose hasta la adolescencia [30].

También se ha encontrado que las sacadas son más cortas y menos precisas en niños que en adultos [30] y que el control cognitivo de la ejecución de la sacada, alcanza un nivel de rendimiento similar al de un adulto en torno a los 10 a 12 años de edad [31] [32] [33].

2.1.1. Inserción de publicidad basada en la atención

En [34] se puede encontrar una descripción interesante de la inserción publicitaria basada en la atención. El primer enfoque se llama publicidad lineal, mientras que el segundo se llama no lineal.

La publicidad lineal insertará clips de anuncios relacionados con el contenido en posiciones temporales menos invasivas. En [35], los autores proponen un enfoque de dos pasos. El primer paso tiene como objetivo seleccionar un anuncio relacionado con el contenido actual. En el documento esto se hizo usando minería de texto. El objetivo del segundo paso es encontrar el momento en el que tenga un bajo grado de tiempo espaciotemporal para insertar el anuncio de una manera menos invasiva.

Otro enfoque es el no lineal [36]. En este caso, también hay dos pasos. El primero consiste en encontrar la ubicación correcta en el marco donde se debe insertar el anuncio. Este paso utiliza el mapa de saliencia para ubicar un área cercana a la más interesante. El segundo paso producirá una armonización del color del anuncio para que sea menos invasivo cuando se proyecta en el marco. Finalmente, el anuncio armonizado se proyecta cerca del área más interesante. Esto conducirá a un anuncio muy notable (cerca de las regiones salientes), mientras que la armonización del color reduce su intrusión

2.1.2. Marketing basado en la atención y optimización de la comunicación

La optimización de marketing se puede aplicar a una gran cantidad de casos prácticos, como sitios web, publicidad, colocación de productos en supermercados, señalizaciones, colocación de objetos 2D y 3D en galerías, etc. Todos estos casos de aplicaciones pueden beneficiarse de los mapas de atención en sí mismos, pero también de las regiones de comparación de interés y posterior análisis de los mapas de atención. Además, la atención sólo indica si las personas notarán el mensaje importante en un anuncio, pero no si lo recuerdan. Por lo tanto, la “memorización” de una imagen es un tema importante donde los modelos de atención pueden ayudar.

Atraer la atención del cliente es el primer paso del proceso de atraer el interés de las personas, inducir el deseo y la necesidad del producto y, finalmente, llevar al cliente a la acción como se describe en la pirámide de AIDA [37]. Es importante enfatizar el hecho de que la atención por sí sola no es suficiente para impulsar a un cliente potencial a la acción, pero al menos es un paso clave hacia esta meta.

Ya hay dos técnicas principales que pueden proporcionar información sobre la atención de las personas en el material de marketing. El primero utiliza estudios de seguimiento ocular (*Eye - Tracking*) en material de marketing. Este enfoque es muy preciso ya que se mide la ubicación precisa de los usuarios en un sitio web/anuncio. Los inconvenientes de este enfoque se encuentran en el tiempo necesario para realizar el estudio, el precio y el hecho de que sólo se pueden probar los documentos terminados o casi terminados.

Otra técnica utiliza el rastreo del *mouse* y también se usa en material de marketing. Hay dos formas de usar esta técnica: ya sea sin una indicación especial o diciéndole a las personas que ubiquen su puntero donde ven [38]. La segunda versión es precisa pero sólo se utiliza para fines de investigación y proporciona una buena aproximación de la mirada del usuario [39] (más del 80 % del seguimiento ocular). En el caso de no dar alguna indicación especial en el seguimiento del puntero, la precisión es mucho menor. La ventaja de las técnicas de rastreo con el *mouse* es que son más baratas que el rastreo ocular, y hay más usuarios disponibles a través de sitios web, mientras que el rastreo ocular necesita que el usuario esté físicamente presente frente a un dispositivo de *Eye - Tracking*. El tiempo de estudio varía, pero podría ser un poco más corto que con el seguimiento ocular.

Finalmente, el método predictivo que es el enfoque principal aquí utiliza mapas de saliencia automáticos. Este enfoque es mucho más rápido que las pruebas de seguimiento ocular (segundos contra días) y también mucho más barato (alrededor de diez veces más barato). Los resultados no son tan buenos como el seguimiento ocular. Los métodos predictivos se pueden lograr en tiempo real y se pueden usar en cualquier momento en el proceso creativo: mientras que los seres humanos (tanto con el seguimiento de los ojos como con el mouse) se verán perturbados por documentos o páginas web sin terminar, un algoritmo automático no lo hará.

Lo anterior significa que el seguimiento ocular o el seguimiento del *mouse* se pueden lograr principalmente una vez que el documento está casi finalizado (lo que podría ser demasiado

tarde para realizar cambios importantes), mientras que el uso de algoritmos automáticos se puede usar en tiempo real y proporciona varios ciclos de retroalimentación durante el proceso creativo. Otra ventaja del método predictivo es que se basa únicamente en imágenes y es posible detectar cualquier tipo de sitio web o anuncio, incluidos los de compañías concurrentes.

2.1.2.1. Compañías basadas en algoritmos de saliencia

Como el enfoque de saliencia y *marketing* es uno de los más cercanos al mercado, se crearon varias compañías basadas en el uso de saliencia en la comercialización.

Feng-GUI [40] es una compañía israelí que se enfoca principalmente en páginas web y optimización de publicidad, incluso el algoritmo también es capaz de analizar secuencias de video. Entre las funciones de abajo hacia arriba (*Bottom - up*) que utilizan, se pueden encontrar color, orientación, densidad y contraste, intensidad, tamaño y peso e intersecciones. Las funciones de arriba a abajo (Top - Down) son detección de rostro, detección de texto y detección de piel. También utilizan información de contexto sobre el tipo de documento (imagen natural, sitio web, cartelera, publicidad) que probablemente corresponda a diferentes densidades de probabilidad según el tipo de documento como en [38]. Las aplicaciones dirigidas principales son las páginas web y la optimización de la publicidad.

3M VAS [41] es el único gran jugador internacional en este campo. Se dan muy pocos detalles sobre el algoritmo utilizado, pero también es capaz de proporcionar visibilidad de video. La principal diferencia con los otros competidores está en los segmentos de clientes con una gama mucho más amplia de posibles aplicaciones. Proporcionan mapas de atención para la optimización de páginas web, pero también anuncios con imágenes o videos estáticos, paquetes o comercialización en la tienda.

EyeQuant [42] es una empresa alemana especializada en la optimización de sitios web. Su algoritmo utiliza pruebas exhaustivas de seguimiento ocular para entrenar el algoritmo y acercarlo al seguimiento visual real para una tarea determinada. También pueden modificar el mapa de saliencia si el espectador está involucrado o no en una tarea o si simplemente pasa por la página modificando sus pesos de algoritmo. Finalmente, proporcionan una pista sobre la “claridad visual” que parece estar relacionada con un estudio sobre la entropía de la imagen. Otros recién llegados como Ittention [43] están entrando en este mercado que muestra un interés creciente en el tema.

Todas esas compañías reclaman alrededor del 90 % de precisión durante los primeros 3 - 5 segundos de visualización [41]. Basan su afirmación en diferentes comparaciones entre su algoritmo y varias bases de datos existentes utilizando varias métricas ROC. Siempre comparan los resultados con la máxima puntuación ROC obtenida por los usuarios humanos. Sin embargo, para imágenes de la vida real y para determinadas tareas y comunicaciones basadas en la emoción, esta precisión disminuye drásticamente, pero sigue siendo útil. Además de esas cuatro compañías, una compañía estadounidense llamada EyePredict [44] propone otro enfoque para el uso de modelos de saliencia. La idea principal es probar un máximo de una combinación de catálogo de productos y proponer la configuración que optimice mejor la visibilidad de un producto dado.

2.1.2.2. Sustitutos comerciales

Se encuentran muchos sustitutos comerciales para las empresas que funcionan en base a los modelos predictivos entrenados con data real de seguimiento ocular. Estos sustitutos se fundamentan en estudios que muestran una fuerte correlación entre la posición del *mouse* y la atención del ojo. En cuanto a lo anterior se encuentran herramientas como mapas de calor basados en los clicks, seguimientos a la navegación del *mouse* mediante el movimiento de este.

En la Tabla 2.1 se observa un *Benchmarking* de este tipo de empresas. Una importante señal es que ninguna de ellas se encuentra en América del Sur.

Compañía	País	\$ Mes * Anual	Click Heatmap	Scroll Heatmap	Dashboard	Form Analytics	Recording	Surveys
Attensee	Alemania	\$49	X	-	-	-	X	X
Crazyegg	EEUU	\$9 - \$99*	X	X	-	-	-	-
Sentialabs	Chile	Custom	X	-	-	-	X	-
Clicktale	EEUU	Custom	X	X	X	X	X	-
Heatmap.ME	EEUU	\$100 - \$200	X	-	X	-	-	-
Hotjar	Malta	\$89 - \$989	X	X	X	X	X	X
Seevolution	EEUU	\$49 - \$299	X	X	X	-	X	-
Ptengine	Japón	\$9 - \$99	X	X	X	-	X	-
Luckyorange	EEUU	\$10 - \$100	X	X	X	X	X	X
Inspetclat	EEUU	\$39 - \$299	X	X	X	X	X	-
Mousestats	Canada	\$29 - \$299	X	X	X	X	X	X

Tabla 2.1: *Benchmarking* empresas sustitutas comerciales

CAPITULO 3: MARCO TEÓRICO

3.1. Pruebas de hipótesis

A menudo uno de los objetivos de la estadística es hacer inferencias acerca de parámetros poblacionales desconocidos con base en información contenida en datos muestrales. Estas inferencias se interpretan de dos formas: como estimaciones de los parámetros respectivos o como pruebas de hipótesis acerca de sus valores. Según la definición de [45], los métodos paramétricos son aquellos en donde la(s) distribución(es) de la(s) cual(es) se toman las muestras están especificadas excepto para los valores de un número finito de parámetros. Los métodos no paramétricos aplican en todos los otros casos. En muchos aspectos, el procedimiento formal para pruebas de hipótesis es semejante al método científico. Este observa la naturaleza, formula una teoría y la confronta con lo observado.

En este contexto, el científico plantea una hipótesis respecto a uno o más parámetros poblacionales de que son iguales a valores especificados. En seguida toma una muestra de la población y compara sus observaciones con la hipótesis. Si las observaciones no concuerdan con la hipótesis, las rechaza. De lo contrario, concluye que la hipótesis es verdadera o que la muestra no detectó la diferencia entre los valores real e hipotético de los parámetros poblacionales. Por ejemplo, un investigador médico puede plantear la hipótesis de que un nuevo medicamento es más eficaz que otro para combatir una enfermedad. Para probarla, selecciona aleatoriamente pacientes infectados con la enfermedad y los divide en dos grupos. El nuevo medicamento A se administra a los pacientes del primer grupo y el B a los del segundo. Entonces, con base en el número de pacientes de cada grupo que se recuperen de la enfermedad, el investigador decide si el nuevo medicamento es más eficaz que el anterior.

Las pruebas de hipótesis se llevan a cabo en todos los campos en los que la teoría se pueda probar contra observación. Un ingeniero de control de calidad puede plantear la hipótesis de que un nuevo método de ensamble produce sólo 5% de piezas defectuosas. Un educador puede decir que dos métodos de enseñanza de lectura son igualmente eficaces, o un candidato político afirmar que la mayoría de los votantes está a favor de él. Todas estas hipótesis pueden ser tema de verificación estadística mediante el uso de datos muestrales observados. ¿Cuál es el papel de la estadística en pruebas de hipótesis? ¿Para qué sirve la estadística en este procedimiento de pruebas de hipótesis?

Probar una hipótesis requiere tomar una decisión cuando se compara la muestra observada contra la teoría. ¿Cómo se decide si la muestra no concuerda con la hipótesis del científico?,

¿cuándo se debe rechazar la hipótesis?, ¿cuándo se debe aceptar? y ¿cuándo no revelar el juicio?, ¿cuál es la probabilidad de que se tome una mala decisión y, en consecuencia, sufrir una pérdida? Y, en particular, ¿qué función de las mediciones muestrales debe emplearse para llegar a una decisión? Las respuestas a estas preguntas están contenidas en un estudio de pruebas de hipótesis estadísticas.

Muchas veces, el objetivo de una prueba estadística es probar una hipótesis concerniente a los valores de uno o más parámetros poblacionales. Por lo general se tiene una teoría, es decir, una hipótesis de investigación, acerca del o los parámetros que se desea apoyar. Por ejemplo, un candidato, Jones, dice que el ganará más de 50 % de los votos en una elección urbana y por tanto saldrá como ganador. Si no se cree en lo dicho por Jones, se podría buscar apoyar la hipótesis de investigación de que Jones no está siendo favorecido por más de 50 % del electorado.

El apoyo para esta hipótesis de investigación, también llamada hipótesis alternativa, se obtiene mostrando (usando los datos muestrales como evidencia) que lo contrario de la hipótesis alternativa, llamado hipótesis nula, es falso. Entonces, una teoría se comprueba demostrando que no hay evidencia que sustente la teoría opuesta: en cierto sentido, una prueba por contradicción. Como se busca apoyo para la hipótesis alternativa de que lo dicho por Jones es falso, la hipótesis alternativa es que p , la probabilidad de seleccionar un votante que esté a favor de Jones, es menor que 0.5.

Si se puede demostrar que los datos apoyan el rechazo de la hipótesis nula $p = 0.5$ (el valor mínimo necesario para conseguir una mayoría) en favor de la hipótesis alternativa $p < 0.5$, se ha alcanzado el objetivo de investigación. Aun cuando es común hablar de probar una hipótesis nula, el objetivo de investigación suele ser demostrar apoyo para la hipótesis alternativa, si dicho apoyo se justifica. ¿Cómo se usan esos datos para decidir entre la hipótesis nula y la hipótesis alternativa? Suponer que $n = 15$ votantes se seleccionan aleatoriamente de una ciudad y se registra Y , el número que está a favor de Jones.

Si nadie en la muestra está a favor de Jones ($Y = 0$), ¿qué se concluiría acerca de lo dicho por Jones? Si Jones en realidad es favorecido por más de 50 % del electorado, no es imposible observar que $Y = 0$ están a favor de Jones en una muestra de tamaño $n = 15$, pero es altamente improbable. Es mucho más probable que se observe $Y = 0$ si la hipótesis alternativa fuera cierta. Entonces, se rechazaría la hipótesis nula ($p = 0.5$) a favor de la hipótesis alternativa ($p < 0.5$). Si se observa $Y = 1$ (o cualquier valor pequeño de Y), un razonamiento análogo lleva a la misma conclusión. Cualquier prueba de hipótesis estadística funciona exactamente de la misma forma y está compuesta de los mismos elementos esenciales.

Los elementos de una prueba estadística son:

1. Hipótesis nula, H_0
2. Hipótesis alternativa, H_a
3. Estadístico de prueba
4. Región de rechazo

Para el ejemplo, la hipótesis a ser probada, llamada hipótesis nula y denotada por H_0 , es $p = 0.5$. La hipótesis alternativa (o investigación), denotada como H_a , es la hipótesis a ser

aceptada en caso de que H_0 sea rechazada. Por lo general la hipótesis alternativa es la que se quiere comprobar con base en la información contenida en la muestra; así, en el ejemplo, H_a es $p < 0.5$. Las partes esenciales de una prueba estadística son el estadístico de prueba y una región de rechazo asociada.

El estadístico de prueba (al igual que un estimador) es una función de las mediciones muestrales (Y en el ejemplo) en las que la decisión estadística estará basada.

La región de rechazo, que de aquí en adelante estará denotada por RR , especifica los valores del estadístico de prueba para el cual la hipótesis nula ha de ser rechazada a favor de la hipótesis alternativa. Si, para una muestra particular, el valor calculado del estadístico de prueba cae en la región de rechazo RR , se rechaza la hipótesis nula H_0 y se acepta la hipótesis alternativa H_a . Si el valor del estadístico de prueba no cae en la RR , se acepta H_0 . Como ya se indicó antes, para el ejemplo pequeños valores de Y llevarían a rechazar H_0 . Por tanto, una región de rechazo que se podría considerar es el conjunto de todos los valores de Y menores o iguales a 2. Se usará la notación $RR = \{y : y \leq 2\}$, o bien, dicho de una forma más sencilla, $RR = \{y \leq 2\}$ para denotar esta región de rechazo.

Hallar una buena región de rechazo para una prueba estadística es un problema interesante que amerita más atención. Es evidente que pequeños valores de Y , por ejemplo $y \leq k$, son contradictorios para la hipótesis $H_0 : p = 0.5$ pero favorables para la alternativa $H_a : p < 0.5$. Entonces de manera intuitiva se selecciona la región de rechazo como $RR = \{y \leq k\}$. Pero ¿qué valor se debe escoger para k ? En forma más general, se busca algunos criterios objetivos para decidir cual valor de k especifica una buena región de rechazo de la forma $\{y \leq k\}$. Para cualquier región de rechazo fija (determinada por un valor particular de k), dos tipos de errores se pueden cometer al llegar a una decisión.

Se puede decidir a favor de H_a cuando H_0 es verdadera (error tipo I), o se puede decidir a favor de H_0 cuando H_a es verdadera (error tipo II). Para la encuesta política de Jones, cometer un error tipo I, es decir rechazar $H_0 : p = 0.5$ (y por tanto aceptar $H_a : p < 0.5$) cuando de hecho H_0 es verdadera, significa concluir que Jones perderá cuando en realidad va a ganar. En contraste, cometer un error tipo II significa aceptar $H_0 : p = 0.5$ cuando $p < 0.5$ y concluir que Jones ganará cuando en realidad va a perder. Para casi todas las situaciones reales, las decisiones incorrectas cuestan dinero, prestigio o tiempo e implican una pérdida. Entonces, α y β , las probabilidades de cometer los errores tipo I y II respectivamente, miden los riesgos relacionados con las dos posibles decisiones erróneas que podrían resultar de una prueba estadística. Como tales, proporcionan una forma muy práctica de medir la bondad de una prueba. La probabilidad α también recibe el nombre de significancia o nivel de prueba. Se recomienda tomar pequeños valores de α , típicamente se utiliza 0.05. [45].

Si W es un estadístico de prueba, el valor p , nivel de significancia alcanzado o p -valor, es el nivel más pequeño de significancia α para el cual la información observada indica que la hipótesis nula debe ser rechazada. Cuanto más pequeño sea el valor de p , es más fuerte la evidencia de que la hipótesis nula debe ser rechazada. Si el valor deseado de α es mayor o igual al valor p , la hipótesis nula debe ser rechazada [45].

3.2. Estadística no paramétrica

Los procedimientos estadísticos no paramétricos se aplican no sólo a observaciones que son difíciles de cuantificar, sino que también son particularmente útiles para hacer inferencias en situaciones en las que existe una seria duda acerca de las suposiciones que son la base de la metodología estándar. Por ejemplo, la prueba t [46] para comparar un par de medias basadas en muestras independientes, está basada en la suposición de que ambas poblaciones están distribuidas normalmente con varianzas iguales. El experimentador nunca sabrá si estas suposiciones se cumplen en una situación práctica, pero con frecuencia estará razonablemente seguro de que las desviaciones de las suposiciones serán lo suficientemente pequeñas para que las propiedades del procedimiento estadístico no se alteren.

Esto es, α y β serán más o menos lo que el experimentador piensa que son. Por otra parte, no es raro que el experimentador dude de la validez de una suposición y se pregunte si está usando un procedimiento estadístico válido. A veces esta dificultad se puede salvar si se usa una prueba de estadístico no paramétrico, con lo cual se evita usar un procedimiento estadístico que sea apropiado sólo para un conjunto muy incierto de suposiciones.

A continuación se definen pruebas de hipótesis útiles al momento de estudiar datos que no cumplen los supuestos de los métodos paramétricos.

3.2.1. Prueba de Levene

Existen diferentes pruebas estadísticas que son útiles al momento de comparar la significancia entre similitud o diferencia entre diferentes grupos de datos. Algunos de estas pruebas, como la de Kruskal- Wallis, tienen como requerimiento que los grupos de datos posean la misma varianza, para ello, existe la prueba de Levene la cual permite evaluar esta condición, considerando tres grupos diferentes la hipótesis nula es $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ [47].

$$\text{Estadístico de prueba} = \frac{(N - k) \sum_{i=1}^k N_i (Z_i - Z_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_i)^2}$$

k es el número de diferentes grupos a los que pertenecen los casos muestreados

N es el número total de casos en todos los grupos,

N_i es el número de casos en el grupo i

Y_{ij} es el valor de la variable medida para el j -ésimo caso del i -ésimo grupo

$$Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_i|, & \bar{Y}_i \text{ es la media del } i\text{-ésimo grupo} \\ |Y_{ij} - \tilde{Y}_i|, & \tilde{Y}_i \text{ es la mediana del } i\text{-ésimo grupo} \end{cases}$$

$$Z_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{N_i} Z_{ij} \text{ es la media de } Z_{ij}$$

$$Z_{ij} = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij} \text{ es la media de } Z_{ij} \text{ para el grupo } i$$

3.2.2. Prueba de Shapiro - Wilk

La prueba de Shapiro - Wilk sirve para verificar si una muestra de datos proviene o no de una *distribución normal*. Su hipótesis nula H_0 indica que la muestra proviene de una distribución normalmente distribuida [48]. Esta prueba es conveniente cuando se desea utilizar alguna prueba estadística que requiera que los datos distribuyan de una forma normal, como por ejemplo, la prueba ANOVA.

$$\text{Estadístico de prueba} = W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$x_{(i)}$ número de la i -ésima posición en la muestra (con muestra ordenada de menor a mayor)

$$\bar{x} = (x_1 + \dots + x_n)/n$$

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

$m = (m_1, \dots, m_n)^T$ con m_1, \dots, m_n los valores medios del estadístico ordenado de variables aleatorias independientes e idénticamente distribuidas, muestreadas de distribuciones normales. V es la matriz de covarianzas de ese estadístico de orden

3.2.3. Prueba U Mann - Whitney

Es una prueba no paramétrica aplicada a dos muestras de datos independientes. Su hipótesis nula H_0 indica que las distribuciones de ambas muestras son iguales mientras que la hipótesis alternativa dice que los valores de una de las muestras tienden a exceder a los de la otra [49].

$$\text{Estadístico de prueba: } U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W$$

n_1, n_2 números de observaciones para muestras I y II respectivamente

W = suma de rangos para la muestra I

Para n_1 y $n_2 > 10$

$$\text{Estadístico de prueba: } Z = \frac{U - (\frac{n_1 n_2}{2})}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

3.2.4. Prueba de Kruskal- Wallis

La prueba de Kruskal - Wallis es una prueba no paramétrica en la cual es posible comprobar si tres o más grupos de datos poseen distribuciones poblacionales idénticas. Considerando tres grupos de datos, la hipótesis nula es H_0 : las tres distribuciones son idénticas, mientras que la hipótesis alternativa dice que al menos dos distribuciones difieren en localización. Esta

prueba no requiere que los datos distribuyan de forma normal pero si es necesario que exista homocedasticidad, es decir, que los grupos posean la misma varianza [50]. Es importante destacar que si se están comparando tres o más grupos de datos, en caso de rechazar la H_0 no se sabe cual de los grupos difiere por lo que es necesario realizar una comparación posthoc entre los grupos para determinar cuales son diferentes.

Estadístico de prueba: $H = \{12/n(n+1)\} \sum_{i=1}^k R_i^2/n_i - 3(n+1)$

n_i = número de mediciones en la muestra de la población i

R_i = suma de rangos para la muestra i , donde el rango de cada medida se calcula de acuerdo con su tamaño relativo en el conjunto general de $n = n_1 + n_2 + \dots + n_k$

3.2.5. Prueba Posthoc de Dunn

Las pruebas de ómnibus unidireccionales, como el análisis de varianza unidireccional común (ANOVA), generalmente plantean hipótesis nulas de que las mediciones en algunos grupos se derivan de una distribución común. Se podría pensar en tales pruebas que respondan preguntas genéricas como: ¿Es necesario mirar más de cerca entre los grupos en busca de diferencias? Sin evidencia para rechazar la hipótesis nula de tales pruebas, el trabajo se pasa a temas nuevos.

Por otro lado, si se rechaza la hipótesis nula de una prueba ómnibus, la pregunta es: ¿Cuál de estos grupos es diferente de cuál? Si se utilizara una prueba ANOVA para probar la diferencia de medias, al rechazar la hipótesis nula, se realizarían comparaciones por pares múltiples utilizando las pruebas de t para la diferencia de medias en datos no emparejados. Sin embargo, la prueba ANOVA tiene supuestos restrictivos con respecto a las distribuciones de los grupos bajo escrutinio: los grupos deben tener varianzas iguales, y las medidas en cada grupo deben ser variables continuas, normalmente distribuidas.

La prueba no paramétrica de Kruskal-Wallis [50] es un análogo no paramétrico a la prueba ANOVA de una vía que sacrifica la precisión de los medios discriminatorios para la discriminación de la dominancia estocástica (es decir, la probabilidad de que una observación extraída al azar de un grupo sea mayor que una observación extraída al azar de otro). Sin embargo, la prueba puede hacerlo independientemente de como se distribuyan las medidas en cada grupo. Si se supone que las medidas son continuas y que las distribuciones no especificadas en cada grupo difieren sólo en su centralidad, entonces se puede entender que la prueba de Kruskal-Wallis es una prueba general para la diferencia de la mediana. Al rechazar la hipótesis nula de esta prueba, se realizarían múltiples comparaciones por pares para la dominancia estocástica o la diferencia de la mediana. La prueba de Dunn es el procedimiento adecuado después de una prueba de Kruskal-Wallis.

Hacer comparaciones de pares múltiples después de una prueba ómnibus redefine el significado de α , que generalmente representa la probabilidad de rechazar falsamente la hipótesis nula para una prueba, dentro del marco inferencial de la prueba de hipótesis. Dunn describió cómo abordar este problema con un ajuste de Bonferroni, que puede modificar el nivel de rechazo para cualquier prueba al dividir α por el número total de pruebas y requiere un valor p mucho menor para rechazar cualquier prueba. Este ajuste deja a α numéricamente intacto,

pero multiplica el valor p . Esto forma la base de la redefinición de la tasa de error familiar (FWER) de α para indicar la probabilidad de rechazar falsamente la hipótesis nula en una prueba de todas las pruebas realizadas.

El ajuste de Bonferroni introdujo el FWER, pero siguieron mejoras adicionales: el ajuste Šidák, que es un enfoque un poco más poderoso pero similar; ajuste secuencial de Holm; y el ajuste secuencial de Holm-Šidák, a veces acreditado a Holland y Copenhaver, que trata las pruebas de hipótesis por pares subsiguientes como partes de diferentes familias sobre la base de si las pruebas anteriores fueron rechazadas. Finalmente, Benjamini y Hochberg razonaron que α debería interpretarse como una tasa de descubrimiento falso (FDR) deseada y debería reflejar cómo cambia la tasa esperada de descubrimientos falsos después de que se rechacen en secuencia algunas pruebas por pares [51] [52].

3.3. Modelo de red neuronal SAM-Resnet

Como fue mencionado en el capítulo de introducción, actualmente el proyecto Akori, en su prototipo de servicio utiliza un modelo de red neuronal para predecir las zonas de mayor atención de las personas en imágenes/páginas webs (saliencias). Dicho modelo es el SAM-Resnet [3].

SAM - Resnet es una extensión de dos populares redes neuronales convolucionales (CNNs), la primera VGG-16 [53] y ResNet-50 [54]. El nombre SAM significa Saliency Attentive Model. En este modelo se propone una nueva arquitectura de predicción de saliencia que incorpora una red convolucional atenta de memoria a corto/largo plazo (Attentive ConvLSTM) que se enfoca iterativamente en ubicaciones espaciales relevantes de una imagen para refinar las características de saliencia. En la arquitectura SAM-Resnet, por primera vez se utiliza un LSTM sin el concepto de tiempo. Este modelo incorpora simultáneamente contextos globales y de escena para inferir la saliencia de la imagen gracias a un LSTM contextual espacial profundo que escanea la imagen tanto horizontal como verticalmente.

La principal novedad de esta propuesta es un modelo convolucional atento, que procesa de forma recurrente las características de saliencia en diferentes ubicaciones, al atender selectivamente las diferentes regiones de un tensor. Por otro lado, esta red puede aprender el sesgo presente en las fijaciones oculares, sin la necesidad de integrar esta información manualmente. Además, según los autores [3], ellos están entre los primeros en incorporar modelos atentos en una arquitectura de predicción de saliencia. A través de una evaluación extensa, la arquitectura propuesta supera el estado del arte actual de en los conjuntos de datos de predicción de saliencia pública. Estos resultados pueden verse en [14]

CAPITULO 4: METODOLOGÍA

Hay dos eventos asociados a las hipótesis de este trabajo de investigación que ocurren actualmente y que no se sabe la respuesta.

Definición 4.1 *Evento 1: EL AUC-Judd de las saliencias entregadas por el modelo SAM-Resnet en las páginas web de eCommerce seleccionadas para el experimento está significativamente por debajo del AUC declarado en el MIT Saliency Benchmark results: cat2000 (AUC-Judd = 0.88, std= 0.04).*

Definición 4.2 *Evento 2: Existen diferencias significativas en las saliencias de los distintos segmentos de edad escogidos en las páginas web de eCommerce mostradas.*

Para calcular la precisión del modelo SAM-Resnet, se pronostica el mapa de saliencias para las imágenes de las webs seleccionadas y luego se utiliza la data obtenida del experimento para generar el mapa de saliencia real (*Ground Truth*) y comparar ambos mediante la métrica *AUC* y correlación de Pearson.

Los resultados para cada uno de los eventos anteriores pueden ser SI o NO de lo cual surgen 4 posibles escenarios los que dependiendo cual de como resultado del experimento guiarán el trabajo de memoria de distintas formas. Los distintos escenarios posibles se observan en la Tabla 4.1.

<p>Evento 1: SI & Evento 2: SI</p> <p>Se usaría el <i>framework</i> propuesto en [26] para adaptar el modelo <i>SAM-Resnet</i> a uno clasificador etario, luego se evalúa el rendimiento para ver si hay mejoras significativas</p>	<p>Evento 1: NO & Evento 2: SI</p> <p>Al igual que el escenario anterior, en este caso, se usaría el <i>framework</i> propuesto en [26] para así adaptar el modelo <i>SAM-Resnet</i> a uno diferenciador por edad</p>
<p>Evento 1: SI & Evento 2: NO</p> <p>Se estudian formas de como construir un modelo adaptado a páginas web de eCommerce y se compara la precisión de este con el modelo de Akori</p>	<p>Evento 1: NO & Evento 2: NO</p> <p>Se discutiría cuales fueron los percances de los resultados, cuales son las mejoras que se proponen para la metodología empleada y se propondrían nuevas líneas de investigación</p>

Tabla 4.1: Escenarios posibles en el transcurso del trabajo de titulo

4.1. Experimento

El experimento consiste en mostrar páginas web de algún rubro de interés para el proyecto Akori a usuarios del rango de edad 18 a 60 años, mientras un dispositivo de *Eye Tracking* guarda la fijación ocular de estas personas. El experimento debe ser en el Web Intelligence Centre y no debe durar más 5 minutos pues los usuarios pierden el interés mirando una pantalla por tiempos más prolongados.

El instrumento de *Eye Tracking* debe ser calibrado para cada persona antes de comenzar con el experimento y la persona debe estar a una distancia promedio de 60 cm de la pantalla. La pantalla del computador en el cual el experimento se realiza es de 19,5”.

4.1.1. Instrumentación

Eye Tracker Tobii Pro x2-30

Eye Tracking Unit	
Sample Rate	30 hz \pm 2
Accuracy	0.4°
Recovery Time	N \A
Mount Type	On Screen, Stand
Screen Size	Up to 25” when mounted (16:9)
Operaing Distance	40-90 cm
Head Movement	50x36
Pupil Size	Yes
Distance	Yes
Connection type	USB 2.0
Data Processing	PC
Weight	0.2 kg
Dimensions	18.4x2.8x2.3 cm

Tabla 4.2: Especificaciones de Hardware Tobii X2-30 [1]

Software Captive Neurolab

Este es el software [55] que se utiliza para ocupar el *Eye Tracker*. A través de este programa se crea un escenario que muestra la secuencia de páginas web a los participantes del experimento. Además, se puede exportar la data en crudo proveniente del seguimiento ocular en un archivo *.csv*, para luego poder aplicar la limpieza de los datos.



(a) Tobii X2-30

(b) Tobii X2-30 Instalado en computador

Figura 4.1: Eye Tracker utilizado en el estudio

4.1.2. Data

Para el tratamiento de la data proveniente del experimento, se usa el software *Python* mediante la plataforma *Colaboratory* [56], debido a que esta última es una plataforma en línea, gratuita y ofrece *GPU* gratis para utilizar las librerías correspondientes en el tratamiento de datos y *Machine Learning* como *numpy*, *pandas*, *OpenCV*, *sklearn*, entre otras. Aquí se hacen las correspondientes transformaciones matriciales para generar los mapas de saliencias y calor respectivamente.

4.1.3. Metodología evaluación SAM-Resnet

Con el fin de evaluar el rendimiento del modelo de red neuronal usado actualmente por el proyecto Akori (SAM-Resnet) en páginas web, es necesario obtener las matrices de saliencias pronosticadas por este en las webs escogidas y luego compararlas con las saliencias obtenidas del experimento. Para lo anterior se instala el sistema operativo Ubuntu y descarga el código asociado al modelo *SAM-Resnet* para utilizar el modelo predictor entrenado.

4.2. Alcances

La cantidad de imágenes a mostrar en el experimento esta acotada debido al tiempo de duración del mismo. A mayor duración, las personas pierden el interés, siendo el promedio de duración de atención en la web de 8 segundos [57] [58], por lo que si se expone cada imagen a los usuarios por un lapso de tiempo mayor muy alejado de este promedio, se podría sesgar la data de seguimiento ocular. Por otro lado, los participantes son más reacios a participar en experimentos voluntarios extensos por disponibilidad de tiempo, por lo que se estima que

la duración total del experimento no debería sobrepasar los 5 minutos.

No se considera la construcción y entrenamiento de modelos de *Machine Learning* en caso de hallar diferencias significativas entre los patrones de saliencias estudiados en esta investigación.

Por otro lado, la muestra de personas del experimento esta restringida al área de Santiago centro y el sector República, esto debido a que es un experimento voluntario en el cual no se dará un pago/beneficio a los participantes, lo que hace más difícil la captación de voluntarios.

El instrumento de *Eye Tracking, Tobii x2-30* está diseñado para ser usado por los usuarios de manera libre y no invasiva, sin tener que utilizar algún tipo de estabilizador de vista. A pesar de lo anterior, existe un margen de error en la precisión del seguimiento ocular debido a que es común que los usuarios no permanezcan completamente estáticos durante la realización del experimento, siendo ideal el uso de un soporte de barbilla para disminuir estos movimientos involuntarios como rotación de cabeza y alejamiento o acercamiento al monitor al mirar. Debido a esto, ciertos mapas de saliencia podrían estar levemente corridos a pesar de haber hecho una correcta calibración del instrumento antes de tomar cada muestra.

Por otro lado, el modelo de *Eye Tracker, Tobii X2-30* no es el dispositivo de seguimiento ocular más preciso y moderno del mercado, por lo que los resultados obtenidos de este experimento son aproximaciones a la realidad.

La duración de esta investigación es de un semestre académico.

4.3. Selección imágenes experimento

4.3.1. Páginas eCommerce

Para el experimento a realizar se escogen páginas web de eCommerce, esto pues según Rocío Ruiz [59] y Magdalena Ortiz [60] parte del *target* de clientes del proyecto Akori son empresas de este rubro, por lo que es relevante considerarlo para el análisis de saliencias de sus webs.

Se escogen 6 sub-rubros distintos de eCommerce, por dos razones, la primera es que se desea tener variedad en el tipo de empresa escogida para poder representar de mejor manera el tipo de páginas webs mostradas en el experimento, de tal manera que los modelos de negocios y el tipo de página web no sean muy similares entre si. La otra razón es que la duración del experimento no debe sobrepasar los 5 minutos pues en caso contrario los participantes pierden la atención y se perdería la información que se quiere captar.

4.4. Segmentos de edad

El principal criterio de selección de los segmentos de edad a estudiar es los potenciales clientes de eCommerce. Por otro lado, dada la complejidad en términos logísticos y legales, se excluye a las personas menores de edad, por lo que los integrantes de cada uno de los segmentos escogidos deben ser mayores de 18 años.

4.5. Pre tratamiento de la data

El primer procedimiento es preprocesar la data cruda del seguimiento ocular por imagen proveniente del *Eye Tracker*. Esta se obtiene desde el software *Captive-NeuroLab*, en donde se exporta un archivo *.csv*, en el cual, entre otros datos, hay columnas correspondientes a la posición x e y de la pantalla en porcentaje en donde miró el usuario en promedio con los dos ojos en el tiempo. Por ejemplo, si el usuario en el tiempo $t = 0$ miró al centro de la pantalla, en el archivo *.csv* se indicará el valor (50 %, 50 %) para las columnas x e y respectivamente.

4.6. Medidas para cuantificar comportamiento visual

A continuación se definen los 3 patrones de comportamiento visual que se cuantifican en el presente estudio. La selección de estos es debido a que existe una metodología propuesta en [26] en donde es posible construir un modelo de saliencias que logra representarlos. De este modo, el modelo puede calibrarse para cada segmento de edad en función de sus diferencias/similitudes en estos patrones.

4.6.1. Exploratividad

Este concepto hace referencia a la dispersión de las zonas observadas de una persona al mirar una imagen por un tiempo determinado.

Para medir la dispersión en la saliencia de una imagen, Krishna et al [26] introducen el concepto de entropía para cuantificar la dispersión del observador en una imagen.

$$H(U_k^g) = \sum_{l \in L} h_{U_k^g}(l) * \log(|L|/h_{U_k^g}(l)), g \in \{g_1, g_2, g_3\} \quad (4.1)$$

U_k^g es el mapa de saliencia del segmento de edad g de la imagen k , $h_{U_k^g}(l)$ es la intensidad del valor del píxel l en el mapa de saliencia U_k^g y L es el conjunto de píxeles de U_k^g .

A mayor entropía, significa que los puntos del mapa de saliencia están mas dispersos lo que se asocia directamente a mayor exploratividad de la escena por parte del observador (ver Figura 4.2).

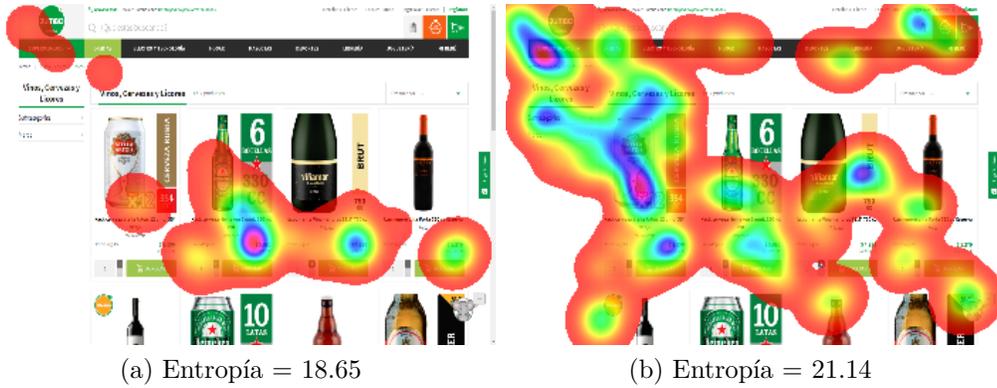


Figura 4.2: Valores de entropía de distintos sujetos para una misma imagen

4.6.2. Acuerdo

El acuerdo o *Agreement* es un concepto que mide si dos mapas de saliencia para una misma imagen cubren las mismas regiones espaciales (zonas de la imagen).

Para cuantificar el acuerdo, la métrica más comúnmente utilizada [61] es el área bajo la curva *AUC* (por su sigla en inglés).

Para lo anterior, dado los segmentos de edad definidos, para cada grupo de personas perteneciente a cada segmento, se consideran los mapas de saliencia de cada imagen del experimento separados por T niveles de probabilidad y se calcula la curva *ROC* la cual grafica la tasa de verdaderos positivos *TPR* como una función de la tasa de falsos positivos *FPR*.

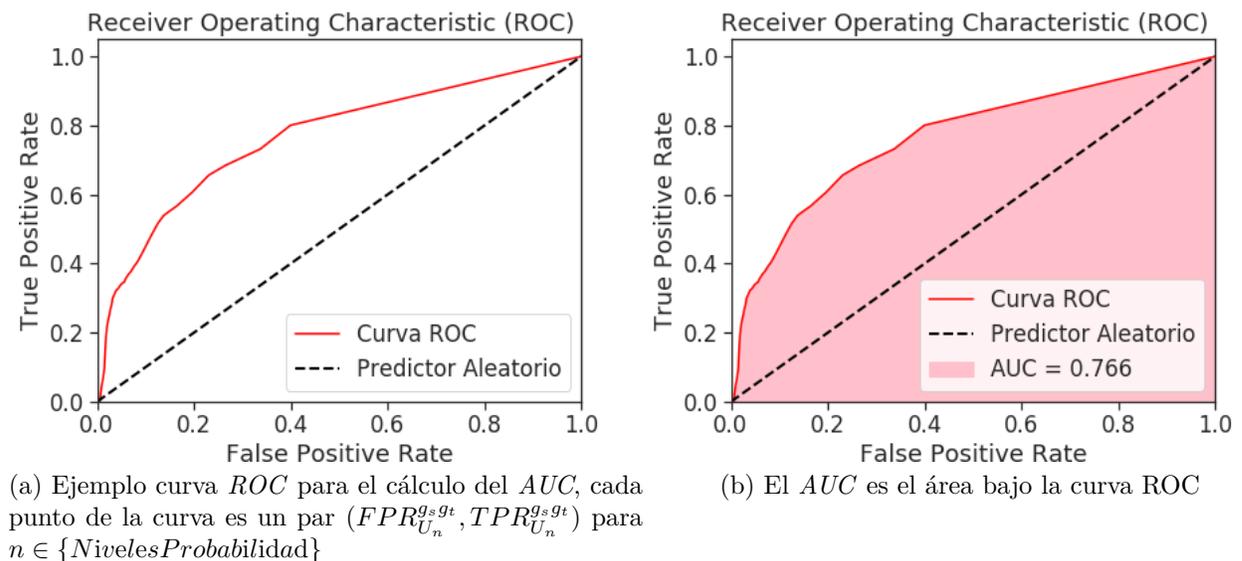


Figura 4.3: Ejemplo curva *ROC* y cálculo de *AUC*

$$TPR_{U_n}^{g_s g_t}(I_i) = \frac{TP_{U_n}^{g_s g_t}(I_i)}{TP_{U_n}^{g_s g_t}(I_i) + FN_{U_n}^{g_s g_t}(I_i)}, s, t \in \{1, 2, 3\} \quad (4.2)$$

$$FPR_{U_n}^{g_s g_t}(I_i) = \frac{FP_{U_n}^{g_s g_t}(I_i)}{TP_{U_n}^{g_s g_t}(I_i) + FN_{U_n}^{g_s g_t}(I_i)}, s, t \in \{1, 2, 3\} \quad (4.3)$$

g_s es el segmento de edad (source) el cual se trata de usar para representar a personas del segmento g_t (target), U_n es el mapa de saliencia que incluye el n % mayor de fijación de las zonas del grupo g_s y g_t respectivamente.

De esta forma se obtienen los $A_{k,n}^{g_i g_j}$, en donde k es la imagen, g_i es el mapa de saliencia del segmento i y n son sujetos del segmento de edad g_j .

4.6.3. Sesgo Central

En el comportamiento visual de las personas observando una imagen, es común hallar una tendencia a mirar en el centro de esta. Esto es debido principalmente a la posición espacial de la imagen dentro de la pantalla y a la configuración de los objetos dentro de la última, los cuales tienden a estar en el centro. Este patrón es conocido como Sesgo Central [62].

En las páginas web la configuración espacial es más variable, en donde existen casos en donde hay más objetos posicionados en el centro y otros en donde se utiliza la mayor parte del espacio dentro de la pantalla.

Para cuantificar el Sesgo central, se utiliza la distancia euclideana entre el centro de masa del mapa de saliencia de una página web en particular y el centro de esta última. Una mayor distancia euclideana significa un menor Sesgo Central y viceversa.

Sean $P1 = (x_1, y_1)$, $P2 = (x_2, y_2)$, las coordenadas de los centros de un mapa de saliencia y de la pagina web respectivamente, la distancia euclideana entre ambos puntos se calcula con la ecuación 4.4.

$$d(P1, P2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4.4)$$

A continuación se define la Correlación de Pearson (CC). El CC no es un patrón dentro de un mapa de saliencia, es una métrica útil para comparar dos mapas de saliencia para una misma imagen. Esta métrica se utiliza en este estudio también de manera independiente a los 3 patrones mencionados anteriormente.

4.6.4. Correlación de Pearson (CC)

El coeficiente de correlación de Pearson r entre dos mapas de saliencia H y P se define como

$$r_{H,P} = \frac{cov(H,P)}{\sigma_H \sigma_P} \quad (4.5)$$

donde $cov(H,P)$ es la covarianza entre H y P y σ_H y σ_P representan las desviaciones estándar de los mapas H y P , respectivamente. El coeficiente de correlación lineal tiene un valor entre -1 y 1. Un valor de 0 indica que no hay una correlación lineal entre los dos mapas. Los valores cercanos a 0 indican una correlación deficiente entre los dos conjuntos. Un valor de 1 indica una correlación perfecta.

Esta métrica sirve para tener una idea si las zonas de mayor saliencia de un mapa están correlacionadas con las zonas de mayor saliencia de otro mapa y lo mismo con las regiones de menor saliencia de ambos mapas.

El signo de r es útil para determinar si los datos comparten la misma estructura. Un valor de -1 también indica una correlación perfecta, pero los datos varían juntos en direcciones opuestas [61].

Para calcular el CC , se reduce la dimensionalidad del mapa de saliencia de dimensión $m \times n$ a un vector de dimensión $1 \times (m \times n)$ como se observa en la ecuación 4.6. De esta manera se puede aplicar la fórmula 4.5 como si H y P fueran dos variables aleatorias de dimensión 1. Esta medida es simétrica y penaliza de igual forma los Falsos Positivos (FP) y Falsos Negativos (FN). Debido a esto, no puede distinguir si las diferencias entre los mapas son debidas a los FP o FN [63].

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & .8 & .7 & .5 & 0 \\ 0 & .6 & .8 & .4 & 0 \\ 0 & .6 & .8 & .3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Reducción}} h = (0, 0, 0, 0, 0, .8, .7, .5, 0, 0, .6, .8, .4, 0, 0, .6, .8, .3, 0, 0, 0, 0, 0) \quad (4.6)$$

4.7. Pruebas de hipótesis

Dependiendo del fenómeno que se está estudiando, existen pruebas de hipótesis que son adecuadas para comparar estadísticamente diferencias o similitudes en las poblaciones de diferentes conjuntos de datos. Algunas pruebas son más exigentes y requieren el cumplimiento de ciertos supuestos.

En esta investigación se comparan patrones en el comportamiento visual para distintos segmentos de edad (cada segmento de edad es un grupo de datos distinto), por lo que una vez cuantificados dichos patrones con métricas, el paso siguiente es comprobar si existen diferencias en estas. Estas diferencias pueden estar medidas por si la media de un conjunto es estadísticamente mayor o menor que la del otro conjunto. Lo anterior es lo realizado en esta investigación.

Las pruebas de hipótesis a usar son las correspondientes a las no paramétricas. Entre ellas Levene, Kruskal - Wallis, Dunn, U Mann Whitney.

CAPITULO 5: EXPERIMENTO Y RESULTADOS

5.1. Procedimiento

La tarea que se le asigna a los participantes del experimento es observar las páginas web de una manera libre, como lo harían de una manera natural, sin restricción. Antes de comenzar con el procedimiento, las personas deben firmar un consentimiento informado en el cual se le da a conocer toda la información importante en cuanto al procedimiento y la confidencialidad de los datos. Los únicos datos demográficos que se almacenan son el sexo, la edad y la ocupación. En Anexos se encuentra una copia del consentimiento informado.

Los sub - rubros escogidos son los siguientes : Retail, e-tail, Viajes, Supermercado, Casa & Hogar, Cupones.

Para cada uno de estos, se escogen las webs de empresas que tienen gran impacto en el eCommerce de Chile:

1. **Retail:** Se escoge *Falabella* debido a que es la empresa de retail chilena que está en el top 5 de las compras online por usuarios chilenos [64], además de ser una de las empresas ganadoras de los eCommerce Awards Chile 2018 [65].
2. **E-tail:** Se escoge *Dafiti* debido a que es una empresa que tiene impacto en las preferencias de los chilenos al momento de comprar en línea y posee el octavo lugar en pedidos en Chile [66]. Además, *Dafiti* está creciendo en Chile: al 2021 pretende triplicar el tamaño de su negocio en el país [67].
3. **Viajes:** Al igual que en la sección retail, *Viajes Falabella* es el líder del eCommerce en Chile [68]. Dado que en la sección retail se escoge *Falabella*, se descarta para esta categoría. Se escoge *Despegar.com* debido a la simpleza de su web en comparación al resto de empresas de aerolíneas.
4. **Supermercado:** Se escoge *Jumbo* al ser el supermercado chileno que le sigue a *Lider* en la cantidad de ventas en línea y además, es el sitio de supermercado que más ha crecido en usuarios únicos visitando su web en el 2017 [69].
5. **Casa & hogar:** En primer lugar esta categoría es escogida debido a que hogar y muebles es la sexta categoría con mayor penetración de vitrineo y compra en Chile. Se escoge *Casa & Ideas* debido a que junto a *Sodimac* son las empresas de este rubro que lideran estas compras. Se decidió no usar *Sodimac* debido a que es perteneciente al

grupo Falabella y ya se escogió una empresa de este holding.

6. **Cupones:** Se escoge *Groupon* debido a que está en el top 3 de compras realizadas por chilenos en la web [64]. Además *Groupon* se ha transformado en una de las mayores empresas de eCommerce en América Latina, luego de la adquisición de sus operaciones en Brasil por el grupo chileno de inversores *Mountain Nazca* [70]

Luego de seleccionar las empresas a estudiar, se escogen 3 imágenes para cada una de sus webs; la primera, corresponde al *Landing Page* debido a que es la imagen de inicio y por lo tanto es la primera impresión, por lo que de esta depende la estadía del usuario en la web. Las otras dos imágenes se escogen de tal manera de que no se repiten aspectos entre las imágenes del resto de las empresas. De esta forma se tienen páginas asociadas a Hombre, Mujer, Tecnología, Gastronomía, Juegos, Disney, Alcohol, Carne, entre otras (ver Figura 8.3 en Anexos).

Por lo anterior, se tienen 6 rubros de 3 páginas cada uno con una duración de 10 segundos más una imagen de paso de 5 segundos entremedio de cada una: $18 \times 10s + 18 \times 5s = 270s = 4,5$ minutos.

Se decide utilizar una imagen negro con un punto blanco al centro como imagen intermediaria de descanso entre las imágenes de páginas web que se muestran en el experimento. El color negro es debido a que este color se caracteriza por tener bajos niveles de luz en la pantalla, lo que provoca que pueda descansar la vista de los observadores. Además, en un experimento similar en donde utilizan un *Eye Tracker* [71] se utiliza una imagen gris como intermediaria entre imágenes mostradas, lo que se asemeja a lo hecho en este estudio.

El punto blanco al centro es un *pop-out* [72], esto es, un objeto para llamar la atención de los usuarios, de tal manera que al presentarse la imagen siguiente, los participantes comiencen a observar libremente la imagen partiendo desde el mismo origen, logrando con esto que cada usuario observe en los primeros segundos lo que realmente más le llama la atención de la página web (ver Figura 5.1).

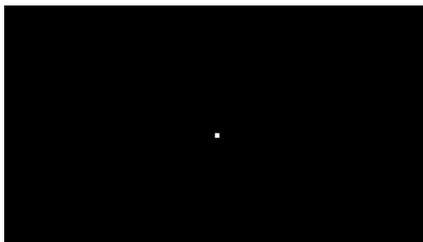


Figura 5.1: Pop-out utilizado en experimento

5.1.1. Selección de participantes

Según una encuesta de Adimark [73] el 46 % de los encuestados en el rango de entre 25 y 34 años de edad compra en línea al menos una vez al mes, de lo anterior surgen tres segmentos interesantes de estudiar, el primero, 18 - 24 años segmento en el cual la mayoría de personas son universitarios no laboralmente activos pero que si están familiarizados con el eCommerce,

el segundo, 25 - 34 años, personas que en su mayoría recibe ingresos y son más activos en compras en línea, finalmente el segmento 35 - 60.

Por la encuesta anterior, por la diferencia generacional y por temas de logística en la captación de participantes en el sector cercano al Web Intelligence Centre (lugar de realización del experimento) es que se definen los 3 grupos de edad mencionados en el párrafo anterior para el análisis en el presente estudio. De esta manera el segmento 1 o S1 es el 18 - 24 años, el segmento 2 o S2 es el 25 - 34 años y el segmento 3 o S3 es el 35 - 60 años.

Se escogió personas saludables para el estudio, es decir que no se les haya diagnosticado alguna enfermedad que pueda interferir en el estudio y que posean una correcta visión.

El método de captación de participantes fue principalmente a través de la plataforma U - Cursos de la universidad de Chile, petición personal a profesores del Departamento de Industrias y funcionarios de la Facultad de Ciencias Físicas y Matemáticas. Además de lo anterior, diariamente el investigador invitó a transeúntes de las calles Domeyko y República a participar del estudio.

Se decide captar un total de 80 personas en el estudio para obtener un conjunto lo más cercano a 20 personas por segmento y no tener problemas con el tamaño de la muestra en el uso de pruebas estadísticas. Además, mientras mayor cantidad de personas, mejor es la representatividad de los datos con la realidad. Por otro lado, en los estudios similares mencionados en el estado del arte, la cantidad de participantes en los experimentos es cercana a 80.

5.1.2. Preprocesamiento datos seguimiento ocular

Luego de exportar en un archivo *.csv* la data de seguimiento ocular del software mediante el cual se realiza el experimento, se eliminan los registros asociados a coordenadas negativas x e y y los valores mayores a 100%, pues, se observó que el instrumento de *Eye Tracking* en algunos casos marca coordenadas afuera de la pantalla y para efectos de este estudio importan sólo las posiciones dentro de la pantalla. Estos casos puntuales pueden deberse al error del instrumento o simplemente a que el usuario miró fuera de la pantalla durante el experimento.

El segundo paso es transformar el sistema de medida de porcentaje a pixeles. Para ello se escogen las dimensiones de los ejes x e y para construir la matriz de fijaciones para la imagen respectiva y luego con esta construir el mapa de saliencia.

Es importante la selección de las dimensiones de la matriz de fijación de un tamaño significativamente menor al de la imagen original mostrada en el experimento, pues, esto ayudará en la reducción de los tiempos de procesamiento en los cálculos matriciales al calcular las medidas de cuantificación del comportamiento visual. La dimensión de las imágenes escogida para la reducción es 240 x 352 pixeles, debido a que es un tamaño considerablemente menor al original y con esta dimensión no se pierde significativamente la calidad de la imagen.

Luego de tener las posiciones de las fijaciones del sujeto en la imagen redimensionada, se aplica un filtro gaussiano (ecuación 5.7) para obtener finalmente el mapa de saliencia y luego aplicando color sobre el mapa de saliencia se obtiene el mapa de calor ver Figura 5.2.

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \quad (5.7)$$

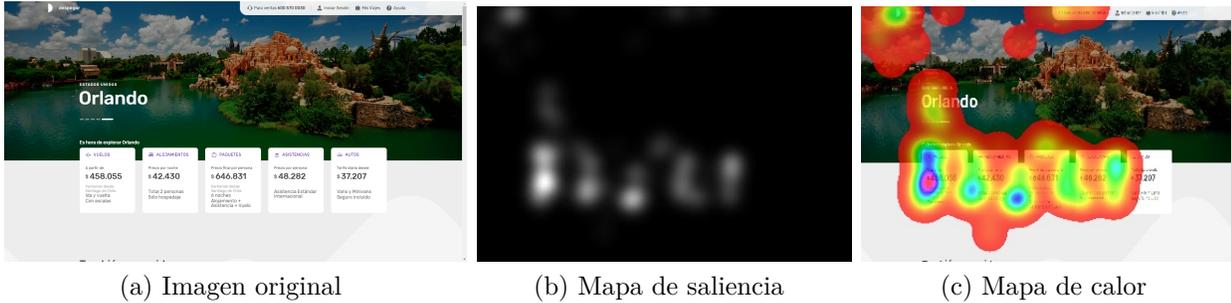


Figura 5.2: Mapa de saliencia vs Mapa de calor

Un mapa de saliencia promedio para un conjunto de sujetos se calcula sumando los mapas de saliencia individuales normalizados por su rango, $S \rightarrow \frac{S-\min(S)}{\max(S)-\min(S)}$, con S un mapa de saliencia y dividido por el numero de sujetos. Luego el mapa de saliencia resultante se vuelve a normalizar por rango obteniéndose el mapa promedio final. Ver Figura 5.3.

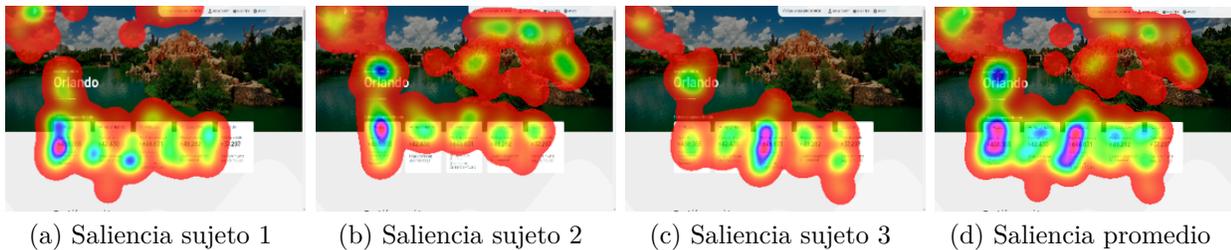


Figura 5.3: Ejemplo mapa de saliencia promedio para una imagen particular

5.2. Análisis de resultados

En esta sección se realiza el análisis de los resultados obtenidos luego de implementar los criterios definidos en la sección anterior en la data obtenida del experimento.

En el análisis estadístico de cada patrón de saliencia estudiado, en cada uno de los enfoques hechos, el primer paso fue observar las distribuciones de los datos y principalmente corroborar normalidad de los conjuntos mediante la prueba de Shapiro – Wilk. Casi en ningún caso los datos distribuyeron de una manera normal (se rechaza H_0 con p-valor <0.05), siendo distribuciones asimétricas en la mayoría de los casos. Debido a esta razón es que las pruebas estadísticas utilizadas son las no paramétricas pues este tipo de pruebas son útiles y arrojan

resultados más confiables en los estudios en donde los datos se comportan diferente a la estadística tradicional.

5.2.1. Grupo experimental

En la tabla 5.1 se puede observar la distribución de edad de los participante del experimento. En anexos Tabla 8.1 se encuentra el detalle de manera desagregada de los datos demográficos de estas personas.

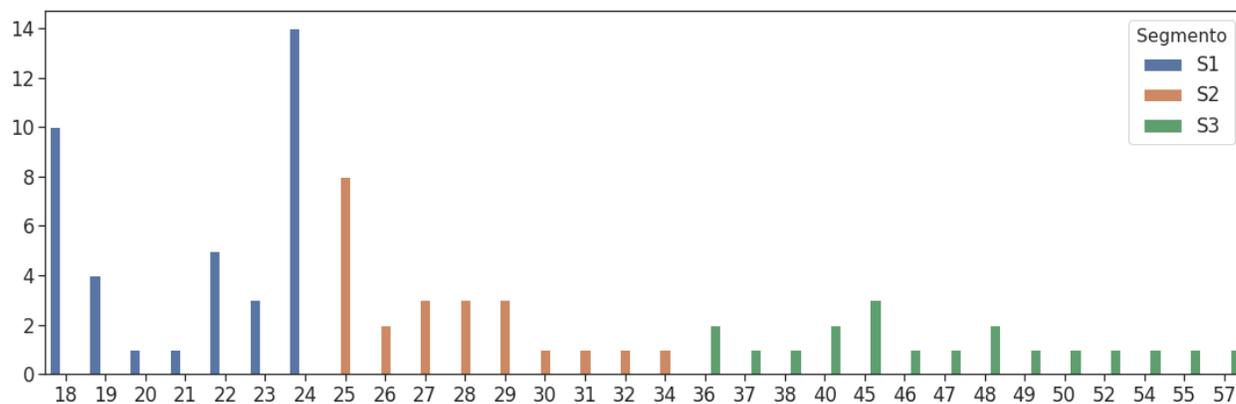


Figura 5.4: Gráfico de distribución de las edades de los participantes del experimento

Segmento	count	mean	std	min	25 %	50 %	75 %	max
S1	38	21.36	2.57	18	18.25	22	24	24
S2	23	27.43	2.55	25	25	27	29	34
S3	19	45.68	6.47	36	40	46	49.5	57

Tabla 5.1: Estadísticas descriptivas para la edad de los participantes del experimento

Los participantes del experimento más jóvenes, fueron en su mayoría estudiantes de Plan Común de Ingeniería, Ingeniería Civil Industrial, Mecánica y Eléctrica.

Para la segmentos de edad mayores, las ocupaciones fueron personal de administración, mantención, garzones, ingenieros, entre otros.

5.2.2. Correlación de Pearson

Se calculan los mapas de saliencia promedio para cada segmento en cada imagen considerando los diez segundos de visualización y se computan las Correlaciones de Pearson para cada combinación de segmentos en cada imagen obteniéndose la Tabla 5.2. Luego se eliminan los casos S1-S1, S2-S2 y S3-S3 pues interesan las combinaciones intersegmentos.

Con esto, cada combinación de segmentos posee 18 valores (asociados a cada imagen) y se calculan las estadísticas descriptivas observadas en la Tabla 5.3a.

	S	T	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
0	S1	S1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	S1	S2	0.93	0.86	0.95	0.94	0.93	0.93	0.95	0.93	0.95	0.94	0.88	0.87	0.93	0.95	0.94	0.85	0.88	0.91
2	S1	S3	0.71	0.66	0.91	0.88	0.86	0.85	0.78	0.66	0.73	0.81	0.80	0.71	0.86	0.78	0.79	0.72	0.74	0.87
3	S2	S1	0.93	0.86	0.95	0.94	0.93	0.93	0.95	0.93	0.95	0.94	0.88	0.87	0.93	0.95	0.94	0.85	0.88	0.91
4	S2	S2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	S2	S3	0.72	0.78	0.92	0.88	0.86	0.86	0.75	0.68	0.77	0.83	0.86	0.75	0.88	0.85	0.83	0.81	0.78	0.82
6	S3	S1	0.71	0.66	0.91	0.88	0.86	0.85	0.78	0.66	0.73	0.81	0.80	0.71	0.86	0.78	0.79	0.72	0.74	0.87
7	S3	S2	0.72	0.78	0.92	0.88	0.86	0.86	0.75	0.68	0.77	0.83	0.86	0.75	0.88	0.85	0.83	0.81	0.78	0.82
8	S3	S3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Tabla 5.2: Correlaciones de Pearson para cada combinación de segmentos en cada imagen

Segmento	count	mean	std	min	25%	50%	75%	max
S1_S2	18	0.92	0.03	0.85	0.89	0.93	0.94	0.95
S2_S3	18	0.81	0.06	0.68	0.77	0.82	0.86	0.92
S1_S3	18	0.78	0.08	0.66	0.72	0.79	0.86	0.91

(a) Estadísticas descriptivas Correlación de Pearson Inter segmentos

Prueba	Estadístico	p-valor
Levene	1.14	0.29
Kruskal-Wallis	1.22	0.26

(b) Pruebas estadísticas para los conjuntos de correlaciones de los S1_S3 y S2_S3

Tabla 5.3: Estadísticas descriptivas CC Inter - Segmentos de edad y resultados de pruebas de hipótesis

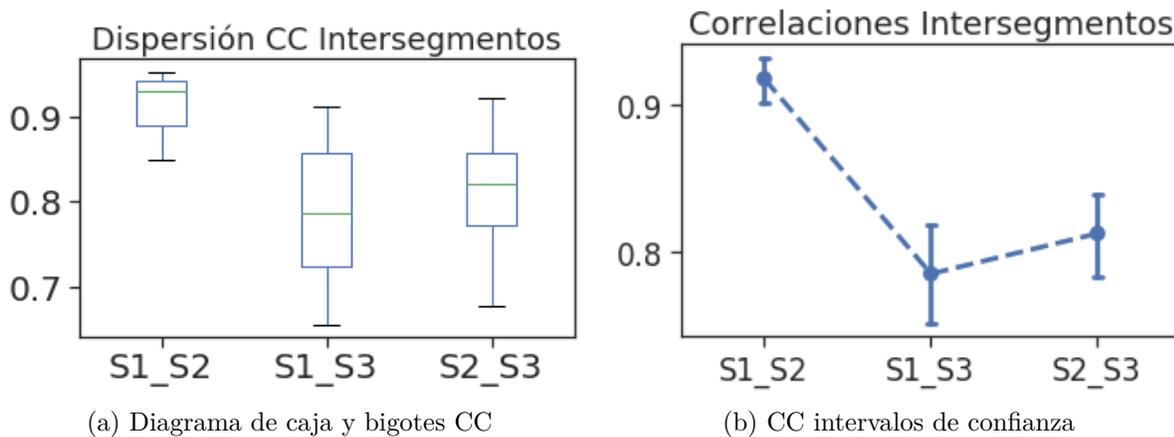


Figura 5.5: Distribución Correlación de Pearson

De los datos se observa que existe una mayor Correlación de Pearson entre el segmento S1 y S2 en comparación a las correlaciones entre las combinaciones de segmentos S1-S3 y S2-S3.

Para corroborar estadísticamente lo anterior, se realiza la prueba de hipótesis de Kruskal-Wallis, para la cual deben cumplirse ciertos supuestos sobre cada grupo de datos a comparar, como por ejemplo, poseer distribuciones similares (asimetría positiva o negativa) y varianzas similares (para esto último es útil la prueba de Levene).

Los resultados de la prueba de Levene para comparar las varianzas de los 3 grupos de edad entregan como resultado que se rechaza la hipótesis nula de que poseen la misma varianza entre estos por lo que no se podría realizar la prueba de Kruskal-Wallis para comprobar las medias de estos 3 conjuntos. Se realiza dicha prueba para comparar los grupos S1-S3 Y S2-S3 ya que en este caso si se cumplen las condiciones necesarias y el resultado es que no se puede rechazar la hipótesis nula de que tienen la misma media ($p - valor > 0.05$) ver Tabla 5.3b.

5.2.3. Exploratividad

Para esta medida, se hacen tres enfoques, en el primero se analiza la entropía de cada mapa de saliencia formado por la visualización de los diez segundos para cada persona en cada imagen, en donde se estudia la diferencia por segmentos de edad de la exploratividad para cada imagen por separado.

En el Enfoque 2 se consideran las entropías de los mapas de saliencia de la visualización de diez segundos de cada persona en cada imagen también pero el análisis estadístico se hace a nivel agregado por segmentos de edad, sin diferenciar por imagen.

En el Enfoque 3 como insumo se utilizan los mapas de saliencia promedio por segmento por imagen. Estos mapas promedio son calculados mediante los mapas de saliencia de la visualización de diez segundos individuales de las personas pertenecientes a cada segmento como en la Figura 5.3.

Enfoque 1

Se realiza la prueba de Levene para los 3 conjuntos de entropía para cada imagen (un conjunto por cada segmento de edad). Dicha prueba es para corroborar si estos conjuntos poseen la misma desviación estándar, pues, esto es un requisito necesario para luego aplicar la prueba de Kruskal - Wallis y con esta última concluir si existen diferencias en los segmentos para esta métrica.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	0.55	1.85	0.09	0.66	1.00	0.50	2.13	0.64	0.80	2.26	0.27	0.16	0.09	0.32	1.36	0.47	1.18	3.01
p-valor	0.58	0.16	0.91	0.52	0.37	0.61	0.13	0.53	0.45	0.11	0.76	0.85	0.91	0.73	0.26	0.62	0.31	0.06

Tabla 5.4: Prueba de Levene conjuntos de entropía por segmento por imagen Enfoque 1

De la Tabla 5.4 se observa que los p -valor para todas las imágenes son mayor a 0.05, por lo que no existe evidencia en contra de que las desviaciones estándar de las entropías para los 3 segmentos en cada imagen son iguales.

El paso siguiente es realizar la prueba de Kruskal - Wallis de manera similar al procedimiento anterior, en donde sólo en tres casos se rechaza la hipótesis nula de que los conjuntos de entropías de los tres segmentos provienen de la misma población (imágenes 1, 4 y 8 , p -valor < 0.05 , ver Tabla 5.5).

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	6.98	5.90	0.94	7.91	2.65	3.91	0.89	13.17	0.85	4.11	2.96	2.08	0.64	1.30	1.75	1.31	2.59	5.35
p-valor	0.03	0.05	0.63	0.02	0.27	0.14	0.64	0.00	0.66	0.13	0.23	0.35	0.72	0.52	0.42	0.52	0.27	0.07

Tabla 5.5: Prueba Kruskal conjuntos de entropía por segmento por imagen Enfoque 1

Finalmente para las imágenes CasaIdeasEstar (I1), DafitiHombre (I4) y DespegarLanding (I8) se realiza la prueba post-hoc de Dunn para ver entre cuales pares de grupos de edad existen diferencias significativas en la entropía de la visualización.

Los resultados de la prueba de Dunn indican que para la imagen 1, existen diferencias significativas entre los segmentos 1 y 3. Para la imagen 4 entre los segmentos 1 y 2 y para la imagen 8 hay diferencias entre los segmentos 1 y 3 y 2 y 3 ($p - valor < 0.05$ en todos los casos), siendo el segmento 3 el estadísticamente menor en todos los casos (Prueba U Mann-Whitney con $p - valor < 0.05$).

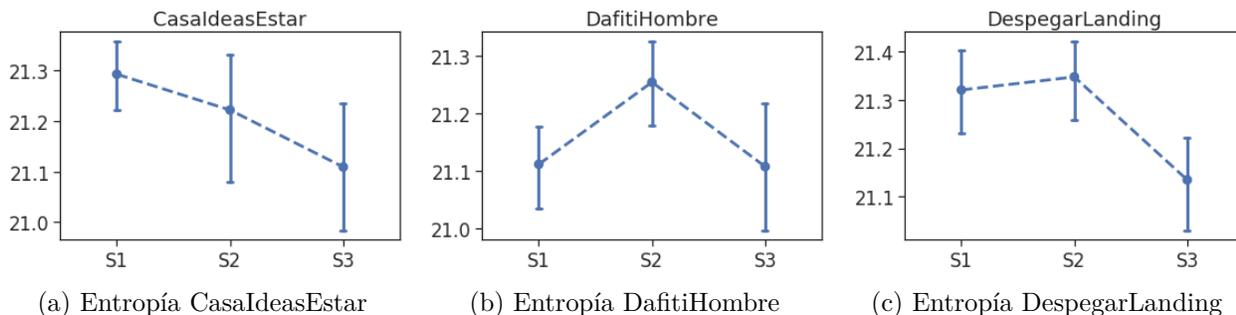


Figura 5.6: Intervalos de confianza para entropía por segmentos en las imágenes que se encuentran diferencias significativas

Adicionalmente se realiza el mismo procedimiento pero considerando los mapas de saliencia de los primeros cinco segundos de visualización. Los resultados son que hay diferencias significativas sólo en 3 de las 18 imágenes (prueba post hoc de Dunn $p - valor < 0.05$ para los casos correspondientes) en donde no se observa una tendencia clara en estas diferencias, pues, en cada uno de los tres casos los resultados son distintos (no hay un segmento en particular que sea estadísticamente mayor/menor para todos estos casos). Debido a lo anterior y a que esto es sólo en un subconjunto pequeño de imágenes es que no se puede concluir que hay diferencias significativas en las entropías de los segmentos de edad estudiados.

Enfoque 2

En este enfoque se tienen 1440 datos correspondientes a 18 entropías por persona (una por cada imagen, $18 \times 80 = 1440$, ver Tabla 5.6a), luego se agrupan las entropías correspondientes a cada segmento y en base a esos conjuntos formados se calculan las estadísticas descriptivas (Tabla 5.6b) y posteriormente pruebas de hipótesis.

	Edad	Entropía	Sexo	Suj	Seg
0	24	20,215	Female	1	S1
1	24	21,105	Female	1	S1
..
1439	29	20,411	Female	80	S2

(a) Matriz entropía Enfoque 2

Seg	count	mean	std	min	25 %	50 %	75 %	max
S1	684	20,30	0,51	18,51	20,02	20,37	20,67	21,51
S2	414	20,38	0,48	18,48	20,09	20,45	20,71	21,72
S3	342	20,19	0,54	17,36	19,89	20,27	20,56	21,31

(b) Estadísticas descriptivas entropía Enfoque 2

Tabla 5.6: Matriz entropía Enfoque 2 y estadísticas descriptivas entropía

Las pruebas de hipótesis indican que no se puede rechazar la hipótesis nula de que los 3 conjuntos tienen la misma varianza (Levene), por lo que se procede a aplicar Kruskal-Wallis en donde se rechaza la hipótesis nula que provienen de la misma población por lo que al menos dos conjuntos difieren (Tabla 5.7a).

Se realiza la prueba posthoc de Dunn, en donde los resultados indican que el segmento 3 difiere del 1 y del 2 y que el segmento 1 con el 2 no difieren ver Figura 5.7a.

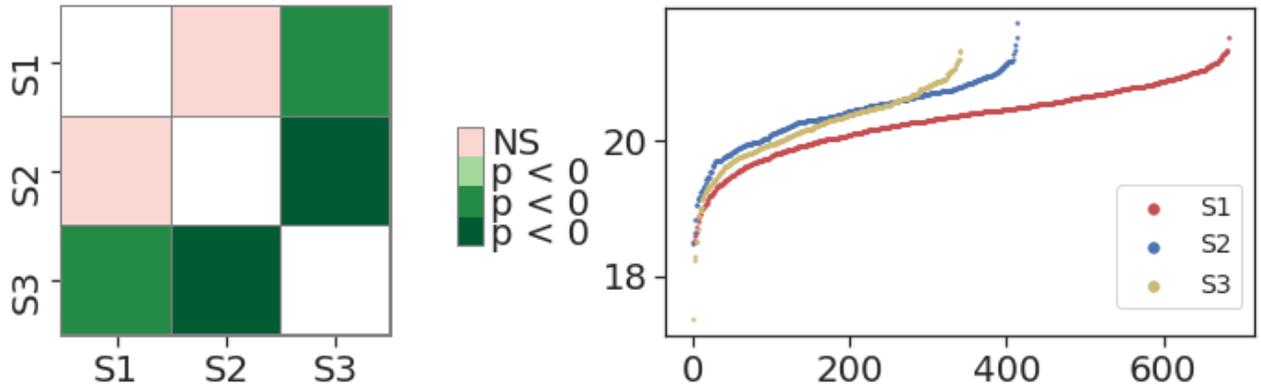
Prueba	Estadístico	p-valor
Levene	2.37	0.09
Kruskal-Wallis	0.23	$7 * 10^{-6}$

(a) Pruebas estadísticas para las entropías de los 3 segmentos Enfoque 2

Prueba	Estadístico	p-valor
Levene	0.77	0.46
Kruskal-Wallis	0.54	0.76

(b) Pruebas estadísticas para las entropías de los 3 segmentos Enfoque 3

Tabla 5.7: Pruebas estadísticas entropía



(a) Resultados prueba de Dunn en donde se observa la significancia en la diferencia de la entropía entre segmentos de edad

(b) Entropía por segmentos ordenada de menor a mayor

Figura 5.7: Entropía Enfoque 1

Enfoque 3

Para este enfoque, se utiliza como insumo los mapas de saliencia promedio por segmento con 100 % de probabilidad. Los resultados son mostrados en la Tabla 5.8. Se utiliza cada fila

S	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
S1	22.15	22.21	22.04	21.45	21.21	21.39	21.69	21.95	21.31	21.67	21.90	21.74	21.62	21.61	21.68	21.97	22.05	21.89
S2	21.96	22.07	21.96	21.61	21.18	21.34	21.61	21.83	21.12	21.59	21.97	21.73	21.61	21.55	21.58	21.95	22.00	22.01
S3	21.98	21.96	21.96	21.50	21.37	21.34	21.62	21.90	21.65	21.65	21.90	21.70	21.63	21.68	21.64	22.03	21.94	22.04

Tabla 5.8: Entropía para mapas de saliencia promedio por segmento para cada imagen Enfoque 2

de la tabla anterior como un vector y en base a estos conjuntos de datos se realizan estadísticas descriptivas (Tabla 5.9) y posteriormente pruebas de hipótesis para analizar significancia en la diferencia o similitud de estos conjuntos de datos.

Las estadísticas descriptivas muestran que casi no hay diferencias en la entropía de los 3 segmentos de edad pues la media, desviación estándar, mediana y valor máximo se diferencia entre los grupos sólo en un par de decimales. En la Figura 5.8a también se puede ver

Segmento	count	mean	std	min	25 %	50 %	75 %	max
S1	18	21.75	0.29	21.21	21.61	21.71	21.96	22.21
S2	18	21.70	0.29	21.12	21.58	21.67	21.96	22.07
S3	18	21.74	0.22	21.34	21.63	21.69	21.95	22.04

Tabla 5.9: Estadísticas descriptivas Entropía para cada segmento Enfoque 2

gráficamente la dispersión de las entropías entre segmentos en donde no se tienen diferencias notorias.

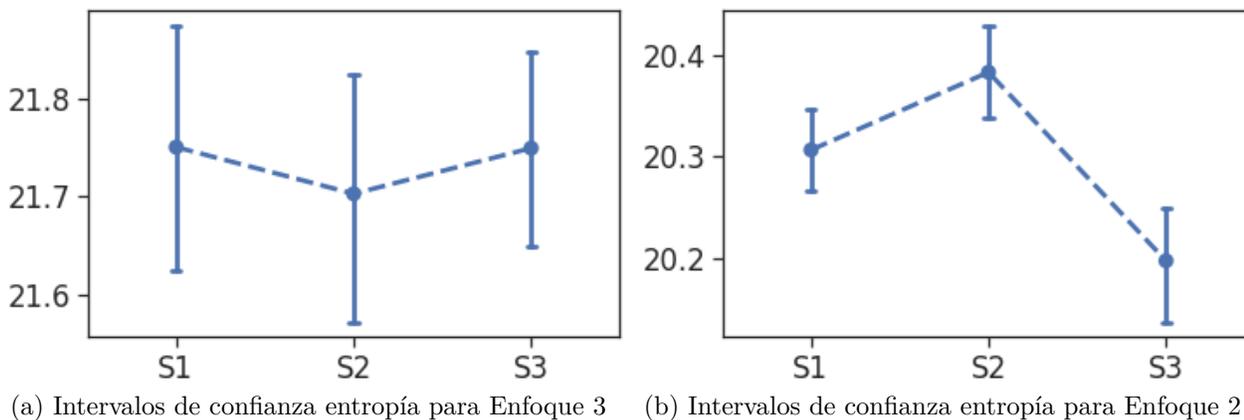


Figura 5.8: Entropía para diferentes enfoques

Adicionalmente, se realizan pruebas estadísticas para corroborar si es que existen diferencias significativas en las entropías de los 3 segmentos. Para esto, se realiza la prueba de Levene en donde no se rechaza la hipótesis nula que estos conjuntos poseen la misma varianza ($p - valor > 0.05$) y a la vez, se observa que cada segmento posee una asimetría positiva en la distribución ($mediana < media$). Esto permite poder realizar la prueba de Kruskal - Wallis en donde el resultado es que no se puede rechazar la hipótesis nula ($p - valor > 0.05$) de que los segmentos poseen la misma distribución de entropía (ver Tabla 5.7b).

5.2.4. Acuerdo

El procedimiento para el cálculo de esta métrica es el siguiente: se computa el mapa saliencia promedio de cada uno de los 3 segmentos para cada imagen y luego se calcula el *AUC-Judd* tomando el mapa promedio como *source* y el mapa de cada sujeto de la imagen del experimento como *target*. De esta forma se obtienen 18 matrices de 3 x 80 (una por cada imagen mostrada en el experimento, ver matriz de ejemplo en la Tabla 5.10) en donde cada componente de la matriz es el *AUC-Judd* de la combinación Segmento-Sujeto en los mapas de saliencia.

Para el cálculo de esta métrica, se realizan dos enfoques, el Enfoque 1 en donde los conjuntos de datos a analizar son los *AUC* promedio de las 18 imágenes por combinación Segmento-sujeto, y el Enfoque 2 en donde se analizan los conjuntos de *AUC* para cada imagen por

Source/Target	1	2	3	4	5	6	7	8	9	10	80
S1	0,82	0,86	0,84	0,92	0,87	0,88	0,80	0,78	0,89	0,89
S2	0,81	0,80	0,82	0,88	0,84	0,90	0,87	0,73	0,90	0,90
S3	0,63	0,79	0,67	0,79	0,66	0,81	0,86	0,69	0,90	0,84

Tabla 5.10: *AUC* en donde el *source* es el mapa saliencia promedio de cada segmento y el *target* es el mapa de saliencia de cada sujeto para una imagen en particular

Source/Target	1	2	3	4	5	6	7	8	9	10	80
S1	0.88	0.83	0.87	0.89	0.83	0.81	0.87	0.86	0.90	0.87
S2	0.86	0.81	0.84	0.85	0.84	0.79	0.90	0.81	0.88	0.89
S3	0.82	0.79	0.76	0.80	0.78	0.80	0.83	0.78	0.88	0.84

Tabla 5.11: Matriz resultante del promedio de las 18 matrices de *AUC*

separado. Para cada enfoque se consideran los mapas de saliencia de los diez segundos de visualización para todas las personas.

Enfoque 1

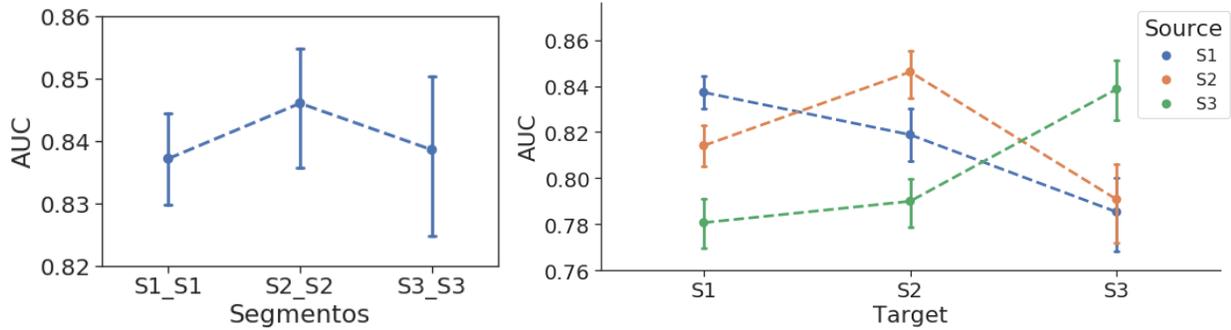
Se promedian las 18 matrices como la del ejemplo Tabla 5.10 para obtener la matriz *AUC* promedio (Tabla 5.11), a partir de esta matriz se seleccionan los sujetos pertenecientes a cada segmento para así formar los conjuntos de *AUC* de las 9 combinaciones Segmento-Target y se calculan las estadísticas descriptivas en base a estos conjuntos de datos (ver Tabla 5.12).

De esta últimas, se observa que el *AUC* es más alto en lo casos Intra-segmentos (S2-S2, S1-S1, S3-S3) que en los casos Inter-segmentos. Esto hace sentido pues los mapas de saliencia promedio provienen de las visualizaciones de las personas que componen cada segmento. Por otro lado, la desviación estándar (*std*) es casi el doble en los casos en que el S3 es el *Target*, lo que demuestra la heterogeneidad en las saliencias de las personas que componen este segmento.

Combi_Segmento	count	mean	std	min	25 %	50 %	75 %	max
S2_S2	23	0.846	0.025	0.787	0.834	0.853	0.862	0.878
S3_S3	19	0.839	0.031	0.753	0.823	0.847	0.864	0.872
S1_S1	38	0.837	0.024	0.772	0.825	0.841	0.853	0.885
S1_S2	23	0.819	0.030	0.757	0.805	0.828	0.838	0.864
S2_S1	38	0.814	0.028	0.750	0.798	0.816	0.836	0.867
S2_S3	19	0.791	0.038	0.681	0.766	0.796	0.814	0.841
S3_S2	23	0.790	0.026	0.736	0.776	0.795	0.804	0.832
S1_S3	19	0.785	0.037	0.679	0.764	0.785	0.810	0.839
S3_S1	38	0.781	0.032	0.719	0.763	0.778	0.803	0.846

Tabla 5.12: Estadísticas descriptivas *AUC* por combinaciones de segmentos

El siguiente paso es realizar pruebas estadísticas para comprobar si existen diferencias



(a) Intervalos de confianza AUC Intra Segmentos (b) Intervalos de confianza AUC Intra e Inter Segmentos

Figura 5.9: Comparación gráfica AUC Intra e Inter segmentos

significativas entre los AUC de un segmento tratando de pronosticar a otro.

Análisis de Acuerdo Intra - Segmentos de edad

La prueba de Levene no demuestra evidencia en contra de que la varianza de los grupos a comparar son distintas (p -valor > 0.05 , ver Tabla 5.13). Esto sumado a la asimetría negativa de estos conjuntos permite realizar la prueba de Kruskal-Wallis, la cual da como resultado que no se puede rechazar la hipótesis nula de que los conjuntos de AUC Intra-Segmentos de edad son de distribuciones diferentes (p -valor > 0.05 , ver Tabla 5.13).

Prueba estadística	Estadístico	p-valor
Levene	0.56	0.57
Kruskal-Wallis	2.44	0.29

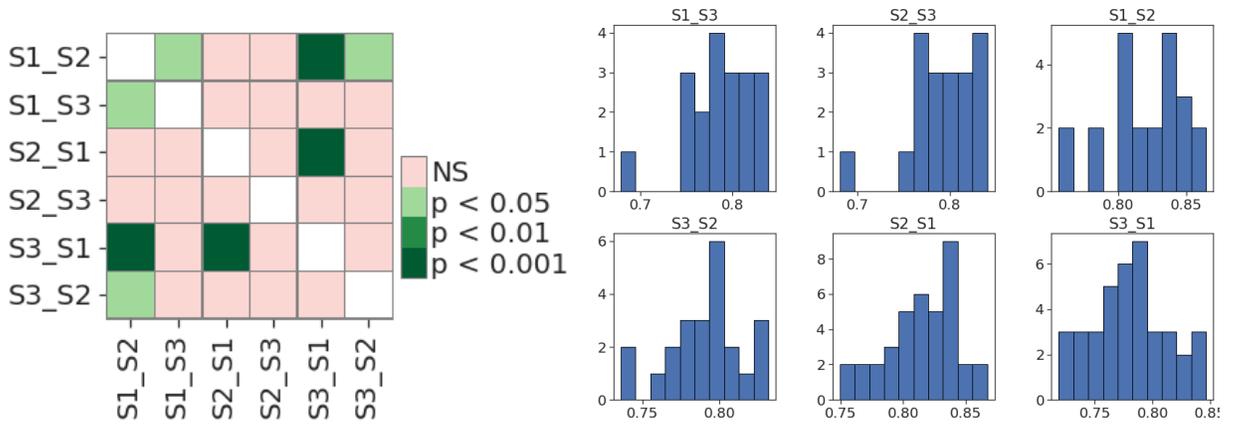
Tabla 5.13: Resultados pruebas estadísticas comparación AUC Intra - Segmentos de edad

Por lo anterior no es posible concluir acerca de que el acuerdo en las fijaciones en algún grupo de edad es más fuerte que el acuerdo de otro grupo.

Análisis de Acuerdo Inter-Segmentos de edad

Al igual que en el caso Intra-Segmentos de edad, se realiza la prueba de Levene y posteriormente Kruskal-Wallis para los conjuntos de AUC de las 6 combinaciones Inter-Segmentos de edad (S1-S3, S2-S3, S1-S2, S3-S2, S2-S1, S3-S1). En este caso, se rechaza la hipótesis nula de que estos conjuntos provienen de la misma distribución (p -valor < 0.05). Esto dice que al menos dos de estas combinaciones Inter-Segmentos difieren, por lo que luego de esto se realiza la prueba de Dunn para comprobar cuales son las combinaciones de grupos que son estadísticamente diferentes.

De la Figura 5.10a los resultados indican que (S1-S3 y S1-S2), (S3-S1 y S1-S2), (S3-S2 y S1-S2), (S3-S1 y S2-S1), son estadísticamente diferentes. De estas 4 combinaciones se



(a) Resultados prueba de Dunn en donde se observa la significancia en la diferencia del acuerdo Inter-Segmentos de edad

(b) Distribución AUC Inter-Segmentos de edad

Figura 5.10: Resultados prueba de Dunn y distribución AUC Inter-Segmentos de edad

observa que el acuerdo entre S1-S2 Y S2-S1 al compararlos con el resto de combinaciones Inter-grupos de edad donde aparece el S3, hay diferencias, siendo estadísticamente menor el acuerdo en las combinaciones Inter-segmentos donde aparece el S3 (prueba de Mann-Whitney, $p - valor < 0.05$). Esto puede deberse a que el acuerdo es mayor (promedio de AUC alto y desviación estándar baja) en los casos en que S1 pronostica a S2 y S2 pronostica a S1 respectivamente. Esto puede verse gráficamente en la Figura 5.9b en donde los intervalos de confianza de S1-S2 y S2-S1 están por sobre los intervalos del resto de combinaciones Inter-Segmentos de edad.

Enfoque 2

Análisis de Acuerdo Intra-Segmentos de edad

De la tabla 5.14 se observa que es posible aplicar la prueba de Kruskal-Wallis para los conjuntos de AUC de las combinaciones S1-S1, S2-S2, S3-S3 ($p - valor > 0.05$ en la prueba de Levene para todas las imágenes).

Sólo se obtiene un $p - valor < 0.05$ en la imagen 9, por lo que se realiza posteriormente la prueba de Dunn la cual da como resultado que existen diferencias significativas entre (S1-S1 y S3-S3) y (S2-S2 y S3-S3), siendo estadísticamente menor el acuerdo Intra Grupo del segmento 3 en comparación a los acuerdos Intra Grupos de los segmentos 1 y 2 para la imagen 9.

Adicionalmente se realiza el mismo procedimiento pero considerando los mapas de saliencia de los primeros cinco segundos de visualización. En este caso, para todas las imágenes, no se encuentran diferencias significativas en el acuerdo Intra - Segmentos de edad. (prueba de Kruskal - Wallis, $p - valor > 0.05$).

Prueba de Levene

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	0.60	1.50	1.05	0.29	0.10	0.27	1.02	0.77	0.11	2.96	3.75	0.16	0.32	0.71	1.14	0.41	0.77	0.93
p-valor	0.55	0.23	0.36	0.75	0.91	0.77	0.36	0.46	0.89	0.06	0.03	0.85	0.73	0.49	0.32	0.67	0.47	0.40

Prueba de Kruskal- Wallis

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	5.36	3.63	1.66	1.26	2.55	0.82	1.40	3.45	11.65	3.53	0.31	0.37	0.03	1.70	2.99	1.20	0.75	1.58
p-valor	0.07	0.16	0.44	0.53	0.28	0.66	0.50	0.18	0.00	0.17	0.86	0.83	0.99	0.43	0.22	0.55	0.69	0.45

Tabla 5.14: Prueba de Levene y Kruskal Wallis *AUC* Enfoque 2 Intra grupos para cada imagen por separado

Análisis de Acuerdo Inter-Segmentos de edad

El procedimiento realizado es similar al anterior pero esta vez considerando las 6 combinaciones Inter-Segmentos de edad (S1-S3, S2-S3, S1-S2, S3-S2, S2-S1, S3-S1).

Prueba de Levene

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	1.08	0.91	0.74	0.49	0.87	0.49	4.20	1.90	3.03	2.84	1.24	0.24	0.53	0.73	1.05	1.71	0.53	2.45
p-valor	0.37	0.48	0.60	0.78	0.51	0.78	0.00	0.10	0.01	0.02	0.30	0.95	0.75	0.60	0.39	0.14	0.75	0.04

Prueba de Kruskal-Wallis

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	30.60	18.46	0.88	12.68	13.29	6.92	NA	43.97	NA	NA	12.29	6.45	2.54	13.87	10.96	5.20	8.92	NA
p-valor	0.00	0.00	0.97	0.03	0.02	0.23	NA	0.00	NA	NA	0.03	0.26	0.77	0.02	0.05	0.39	0.11	NA

Tabla 5.15: Prueba de Levene y Kruskal Wallis *AUC* Enfoque 2 Intra grupos

En los casos que corresponde, se rechaza la hipótesis nula de que los acuerdos Inter Segmentos provienen de la misma distribución para 7 de las 18 imágenes ($p - valor > 0.05$ para prueba Levene y < 0.05 para la prueba de Kruskal-Wallis) por lo que se procede a realizar la prueba de Dunn para analizar diferencias significativas entre el acuerdo Inter Segmentos.

Los resultados de la prueba posthoc de Dunn son los siguientes:

- En 3 de las 7 imágenes existen diferencias significativas entre las combinaciones (S3_S1 Y S1_S2) y (S3_S1 Y S2_S1)
- En 2 imágenes diferentes de las anteriores existen diferencias significativas entre las combinaciones (S3_S2 y S2_S1) y (S3_S1 y S2_S1)
- En las 2 imágenes restantes no existen diferencias significativas entre ninguna combinación de Inter Segmentos

Repetiendo el procedimiento anterior pero considerando los mapas de saliencia de los primeros cinco segundos de visualización, se obtiene que para cinco de las 18 imágenes existen diferencias significativas en el acuerdo Inter - Segmentos de edad (prueba posthoc de Dunn, $p - valor < 0.05$), en donde el acuerdo S3_S1 es estadísticamente menor al S1_S2 y S2_S1 (prueba U Mann - Whitney, $p - valor < 0.05$).

5.2.5. Sesgo Central

Como fue explicado previamente, el patrón de tendencia al centro es cuantificado a través de la distancia euclidiana en píxeles entre el centro de masa del mapa de saliencia y el centro de la página web.

Se calculan las pruebas estadísticas de Levene y Kruskal-Wallis respectivamente para los conjuntos de distancias euclidianas para los 3 segmentos de edad en cada imagen. Los resultados se observan en la Tabla 5.16.

Prueba de Levene																		
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	0.72	2.08	1.50	0.81	0.04	0.65	2.92	0.45	1.34	0.28	0.08	1.70	1.28	0.96	1.13	0.54	3.80	2.09
p-valor	0.49	0.13	0.23	0.45	0.96	0.53	0.06	0.64	0.27	0.76	0.92	0.19	0.29	0.39	0.33	0.59	0.03	0.13
Prueba de Kruskal- Wallis																		
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
Estadístico	1.67	1.07	2.41	0.33	7.16	18.20	7.49	3.10	12.76	4.08	1.82	8.90	0.01	0.03	0.23	0.44	1.22	0.15
p-valor	0.43	0.59	0.30	0.85	0.03	0.00	0.02	0.21	0.00	0.13	0.40	0.01	1.00	0.99	0.89	0.80	NA	0.93

Tabla 5.16: Pruebas estadísticas Sesgo Central

Existen diferencias significativas en el sesgo central entre los segmentos de edad para 5 de las 18 imágenes ($p - valor > 0.05$ para prueba Levene y < 0.05 para la prueba de Kruskal-Wallis).

Se realiza la prueba posthoc de Dunn para ver cuales son los pares de segmentos que son estadísticamente diferentes. Los resultados son los siguientes:

1. DafitiLanding (I5): S1 y S3. S3 es significativamente mayor
2. DafitiMujer (I6): S1 y S3, S2 y S3. S3 es significativamente mayor que S1 y S2
3. DespegarDisney (I7): S2 y S3. S3 es significativamente mayor que S2
4. DespegarOrlando (I9): S1 y S3, S2 y S3. S3 significativamente mayor que S1 y S2
5. FalabellaTech (I12): S1 y S2, S2 y S3. S1 y S3 son significativamente mayores que S2

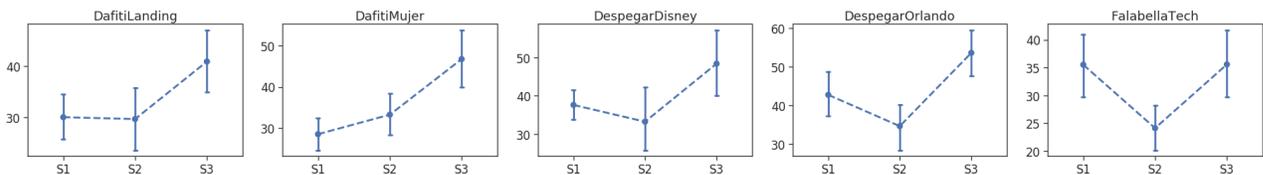


Figura 5.11: Intervalos de confianza distancia euclidiana por segmentos para imágenes 5, 6, 7, 9 y 12 respectivamente

Se realiza el mismo procedimiento pero considerando los mapas de saliencia de los primeros cinco segundos de visualización. En este caso, para todas las imágenes, no se encuentran diferencias significativas en el Sesgo Central (prueba de Kruskal - Wallis, $p - valor > 0.05$).

5.2.6. Mapas de Saliencia imágenes por segmento

A continuación, en las Figuras 5.12, 5.13, 5.14, se muestran los mapas de saliencias promedio por segmento de las imágenes mostradas en el experimento. Estos mapas son considerando los diez segundos de visualización por imagen. Se puede observar para cada uno de estos ejemplos que los segmentos 1 y 2 son casi idénticos para todos los niveles de probabilidad de saliencia mientras que el segmento 3 difiere levemente con los anteriores. Esto es mayormente en las zonas de más probabilidad de observación (colores azul oscuro y morado en el mapa de calor), lo que hace sentido con lo demostrado en los valores de las métricas calculadas anteriormente.

Para efectos prácticos, gráficamente a través de las imágenes se puede observar que las diferencias entre los 3 segmentos no son significativas ya que existe una coincidencia en la mayoría de las zonas dentro de cada imagen entre los segmentos.

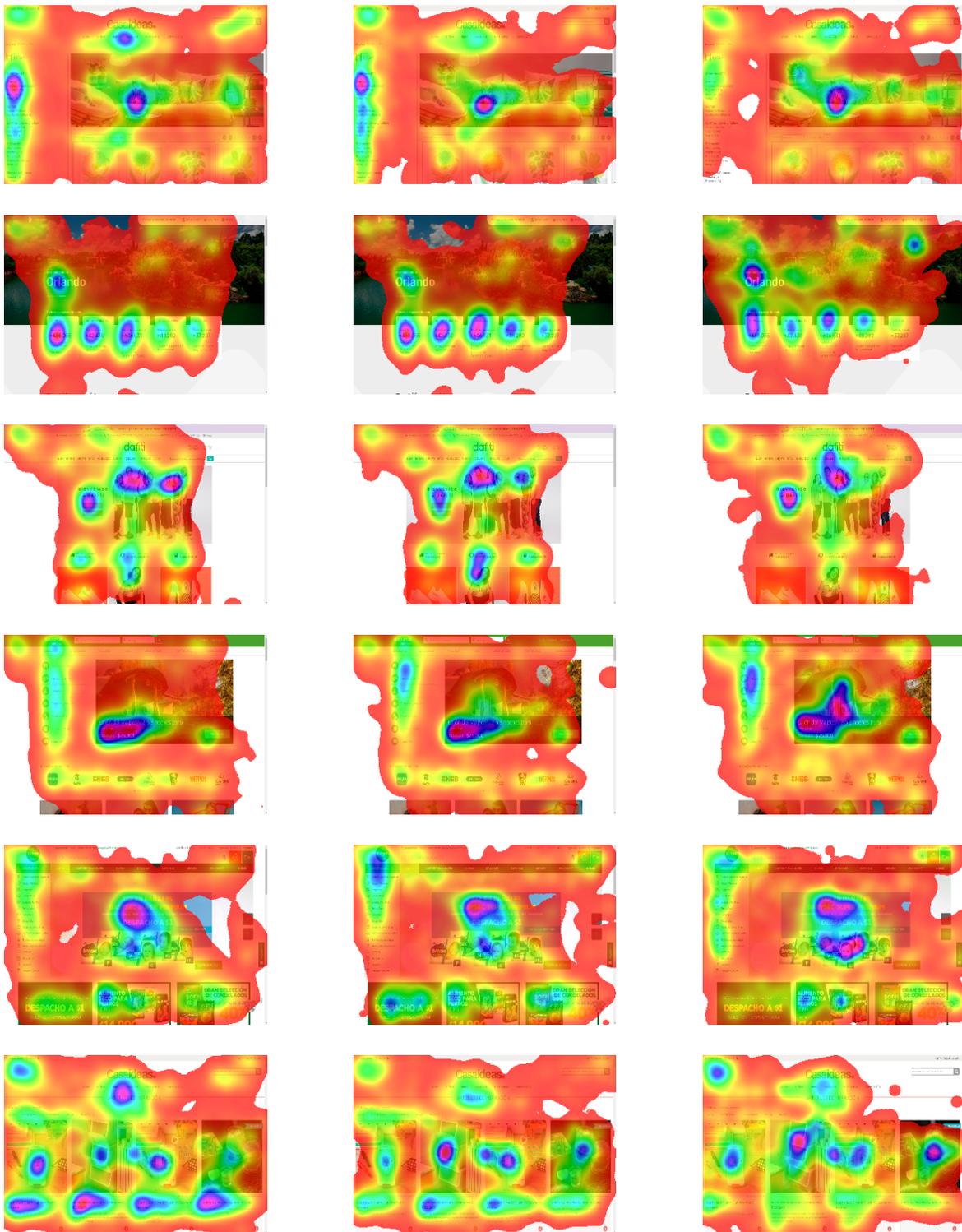


Figura 5.12: Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente

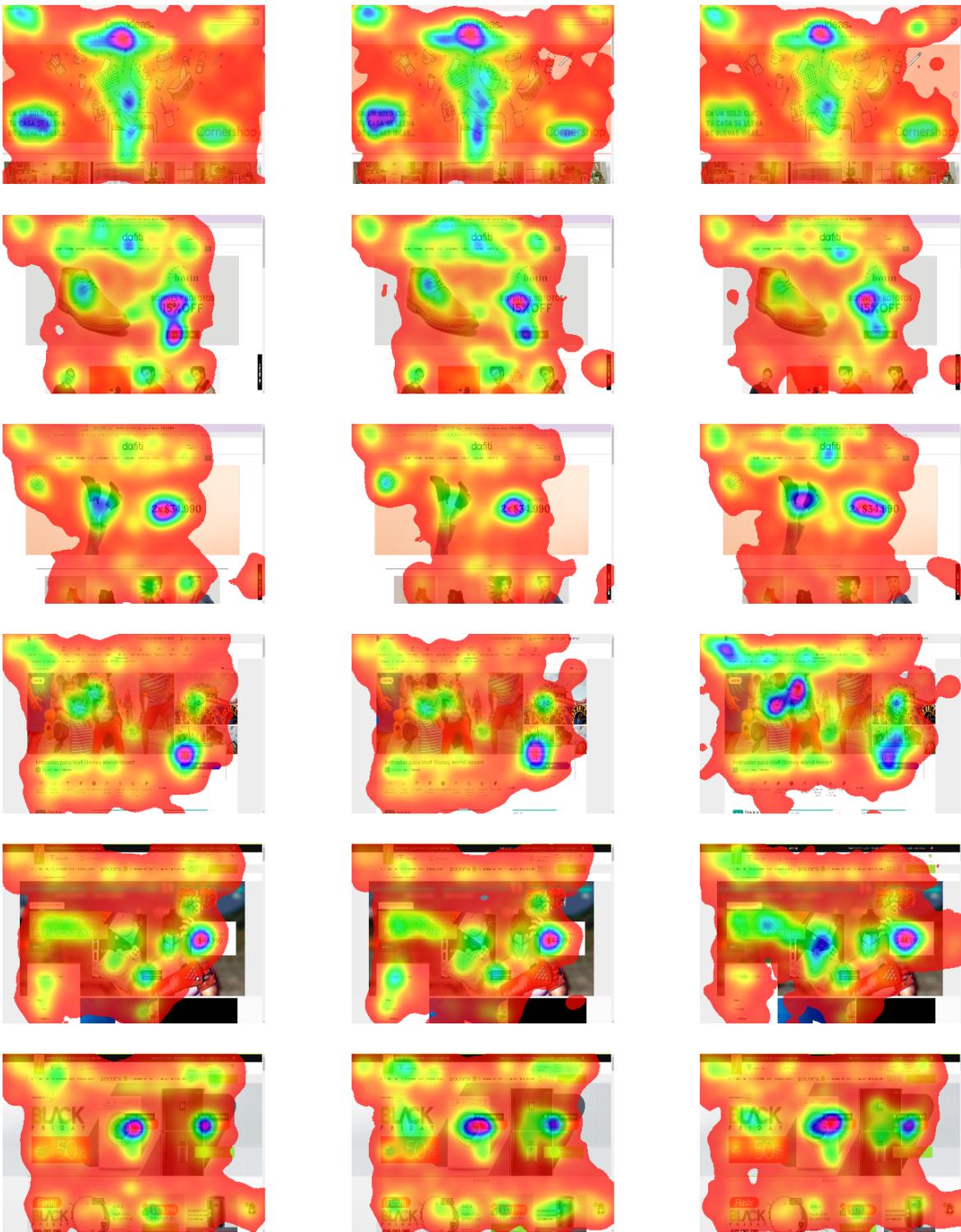


Figura 5.13: Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente

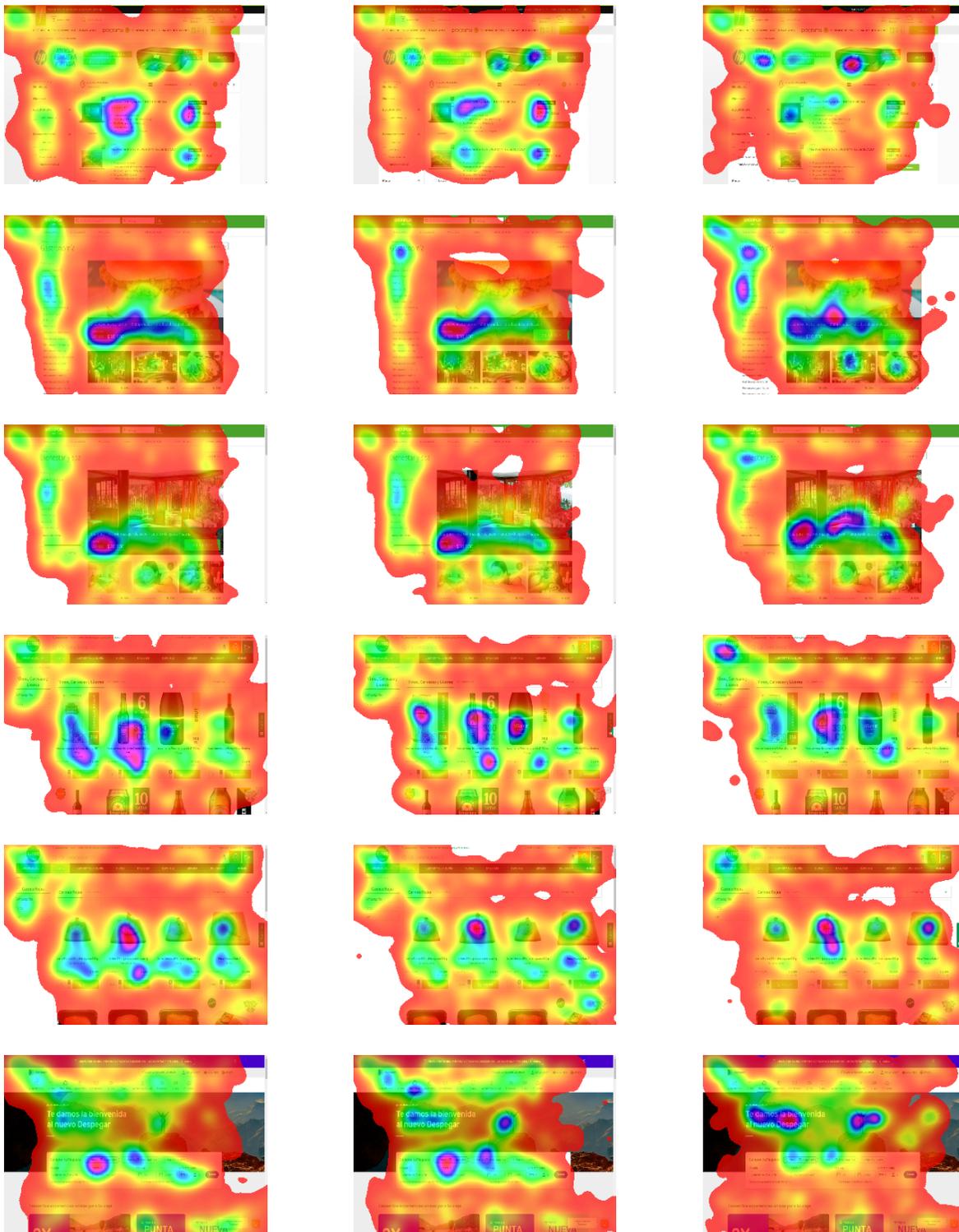


Figura 5.14: Ejemplos saliencias promedio en imágenes por segmentos S1, S2 y S3 respectivamente

5.3. Evaluación modelo de saliencia SAM-Resnet

Para la evaluación del modelo SAM-Resnet se consideran la visualización de todos los participantes del experimento sin segmentar para calcular los mapas de saliencia promedio para cada imagen y luego compararlos con las saliencias pronosticadas por el modelo SAM-Resnet.

El procedimiento mencionado, se aplica para dos conjuntos de páginas webs de eCommerce diferentes. El primero corresponde a las 18 imágenes usadas para el análisis de diferencias por segmentos etarios y el otro conjunto son 15 páginas webs en donde se tiene la visualización de 48 personas para cada una. El experimento del último conjunto fue similar al hecho en este estudio también en el marco del proyecto Akori y fue realizado por el memorista Pablo Uribe.

La evaluación del desempeño del modelo se realiza mediante dos métricas, la primera es el *AUC-Judd* y la segunda es la Correlación de Pearson (CC). Se considera la visualización por diez y cinco segundos en cada imagen.

Evaluación mediante métrica *AUC-Judd*

En la Tabla 5.17 y 5.18 se observan los *AUC* por imagen y en la Tabla 5.19 las estadísticas descriptivas del conjunto de *AUC* formado por todas las imágenes para las visualizaciones por cinco y diez segundos. Este último conjunto distribuye de forma normal para ambos casos (prueba de Shapiro $p - valor > 0.05$) en donde el promedio es 0.75 y la desviación estándar 0.05 para la visualización por diez segundos y 0.78, 0.05 para los cinco segundos. Estos valores son diferentes a los declarados en [14] siendo 0.88 y 0.04 para el promedio y desviación respectivamente. Los resultados obtenidos no son idénticos, sin embargo se considera como un buen nivel de *AUC* el obtenido al lograr acertar bien la mayor parte de las saliencias, siendo levemente más preciso (0.03) el modelo en el caso de que trata de pronosticar los primeros segundos de visualización de las personas.

Estos resultados dan como primero indicios que el modelo SAM-Resnet si bien es bueno para predecir saliencias en páginas web de eCommerce, existen ciertas características en la saliencia de las páginas web que poseen más prioridad sobre otras y que el modelo no logra capturar.

Conjunto de imágenes 1

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
10 seg	0.61	0.63	0.69	0.76	0.80	0.77	0.72	0.73	0.79	0.73	0.70	0.76	0.76	0.77	0.74	0.74	0.72	0.74
5 seg	0.66	0.66	0.75	0.78	0.82	0.8	0.75	0.75	0.79	0.79	0.73	0.8	0.76	0.79	0.75	0.78	0.76	0.8

Tabla 5.17: *AUC-Judd* para conjunto de imágenes 1

Conjunto de imágenes 2

	I19	I20	I21	I22	I23	I24	I25	I26	I27	I28	I29	I30	I31	I32	I33
10 seg	0.81	0.83	0.85	0.74	0.74	0.74	0.74	0.67	0.74	0.88	0.78	0.80	0.73	0.86	0.79
5 seg	0.82	0.85	0.87	0.76	0.68	0.77	0.75	0.78	0.8	0.75	0.89	0.81	0.82	0.75	0.87

Tabla 5.18: *AUC-Judd* para conjunto de imágenes 2

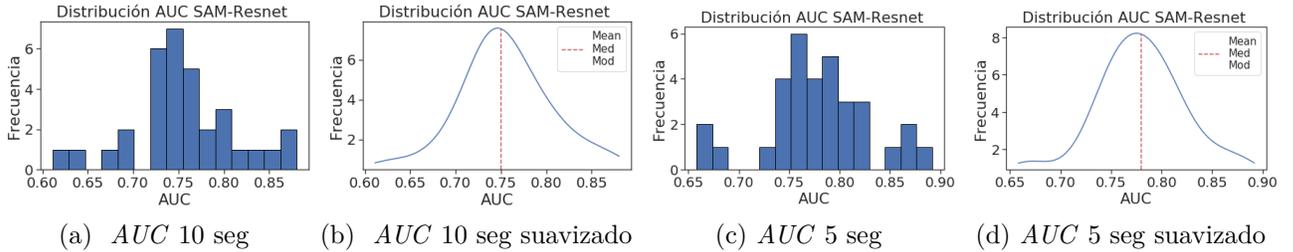


Figura 5.15: Distribución de *AUC* modelo SAM-Resnet para 10 y 5 segundos de visualización considerando los dos conjuntos de imágenes

	count	mean	std	mediana	moda
AUC SAM-Resnet 10 seg	33	0.75	0.05	0.75	0.75
AUC SAM-Resnet 5 seg	33	0.78	0.05	0.78	0.78

Tabla 5.19: Estadísticas descriptivas *AUC* SAM-Resnet para el conjunto de todas las páginas web

Evaluación mediante métrica Correlación de Pearson

Para el caso de la Correlación de Pearson, los resultados de la media y la desviación son 0.61 y 0.13 para el conjunto de todas las páginas web miradas por diez segundos y 0.64 y 0.13 para los 5 segundos. Los valores declarados por el *MIT* son 0.89 y 0.06 para la media y desviación respectivamente.

Por otro lado, para esta métrica, la desviación estándar es alta en los resultados del presente estudio ($std = 0.13$ para ambos casos) y los gráficos de distribución de las Figuras 5.16b y 5.16d muestran una asimetría negativa, lo que implica que hay más datos de correlaciones separados a la izquierda de la media.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
10 seg	0.38	0.32	0.62	0.69	0.74	0.75	0.50	0.64	0.75	0.59	0.66	0.74	0.54	0.77	0.43	0.70	0.72	0.80
5 seg	0.43	0.38	0.72	0.76	0.75	0.73	0.61	0.68	0.7	0.62	0.74	0.82	0.41	0.76	0.39	0.77	0.85	0.87

Tabla 5.20: Correlaciones de Pearson conjunto de imágenes 1

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
10 seg	0.64	0.42	0.65	0.58	0.54	0.46	0.36	0.63	0.46	0.46	0.85	0.69	0.76	0.67	0.62
5 seg	0.65	0.43	0.65	0.6	0.55	0.5	0.35	0.68	0.47	0.47	0.85	0.66	0.75	0.73	0.65

Tabla 5.21: Correlaciones de Pearson conjunto de imágenes 2

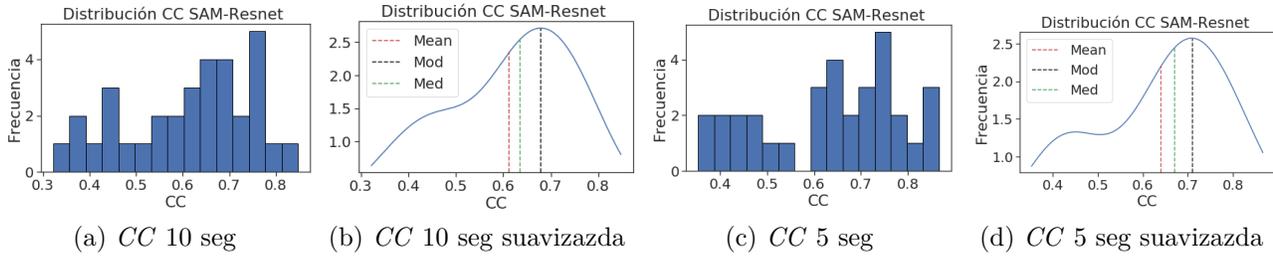


Figura 5.16: Distribución de CC modelo SAM-Resnet para 10 y 5 segundos de visualización considerando los dos conjuntos de imágenes

	count	mean	std	mediana	moda
CC SAM-Resnet 10 seg	33	0.61	0.13	0.64	0.68
CC SAM-Resnet 5 seg	33	0.64	0.13	0.67	0.71

Tabla 5.22: Estadísticas descriptivas CC SAM-Resnet para el conjunto de todas las páginas web

5.3.1. Comparación gráfica de saliencias Ground Truth vs SAM-Resnet

En las Figuras 5.17 y 5.18 por fila, se puede observar las saliencias promedio para una imagen del conjunto 1 de todos los participantes del experimento (Ground Truth) considerando 10 y 5 segundos de visualización vs las saliencias para dicha imagen pronosticadas por el modelo SAM-Resnet.

Lo anterior de manera similar para las imágenes del Conjunto 2 se observa en la Figura 5.19. En este caso, en cada fila hay dos pares de imágenes de la siguiente forma : Ground Truth 10 segundos imagen 1 - Saliencia SAM-Resnet 1, Ground Truth 10 segundos imagen 2 - Saliencia SAM-Resnet 2, y así sucesivamente.

El Ground Truth considerando 5 segundos para el Conjunto 2 de imágenes, al igual que el del Conjunto 1, es muy similar al de 10 segundos. Por dicha razón se decide no incluirlo de manera gráfica.

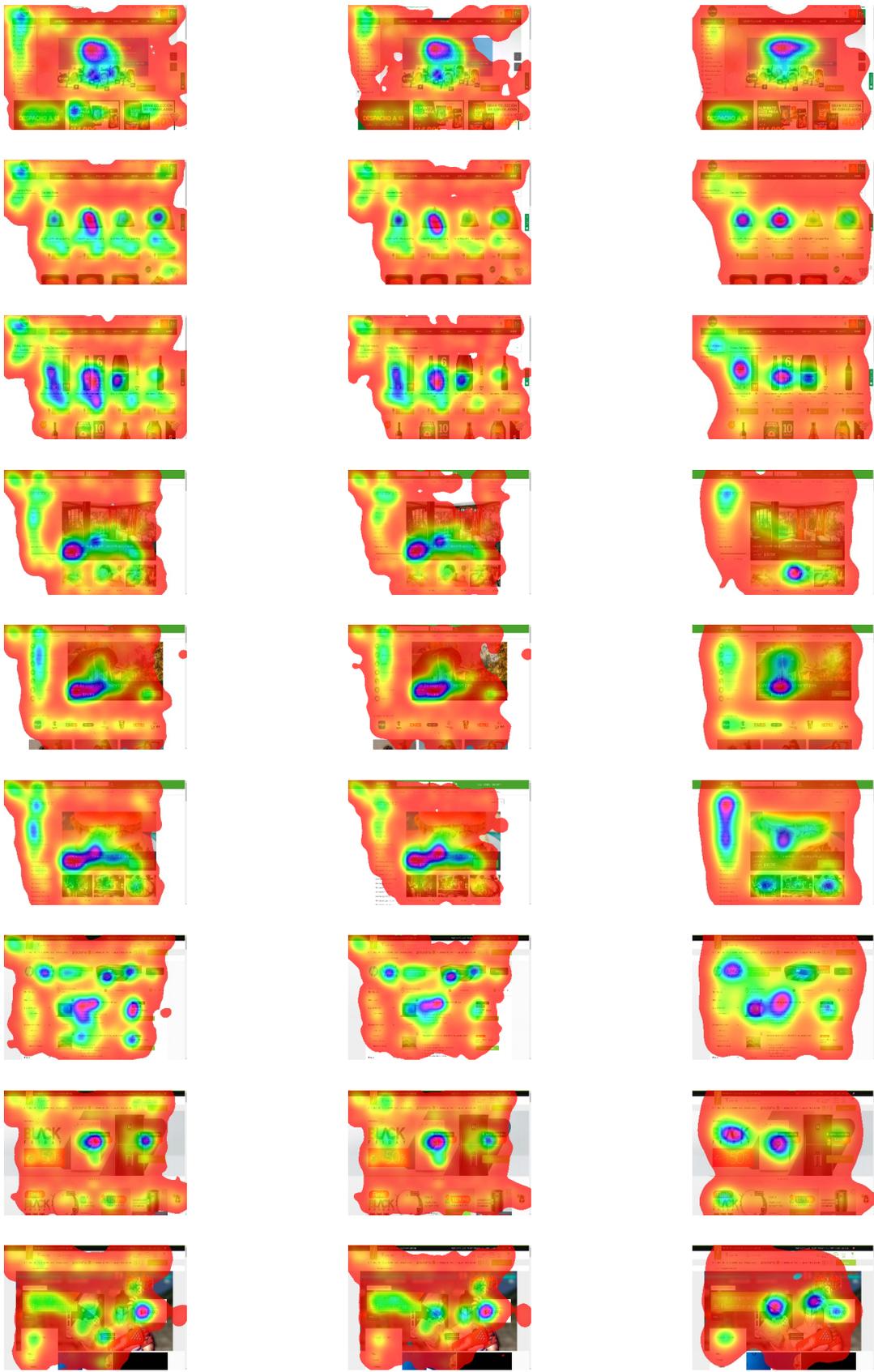


Figura 5.17: Por fila, Ground-Truth 10 y 5 segundos de visualización y saliencia entregada por SAM-Resnet respectivamente para una imagen en particular Conjunto 1

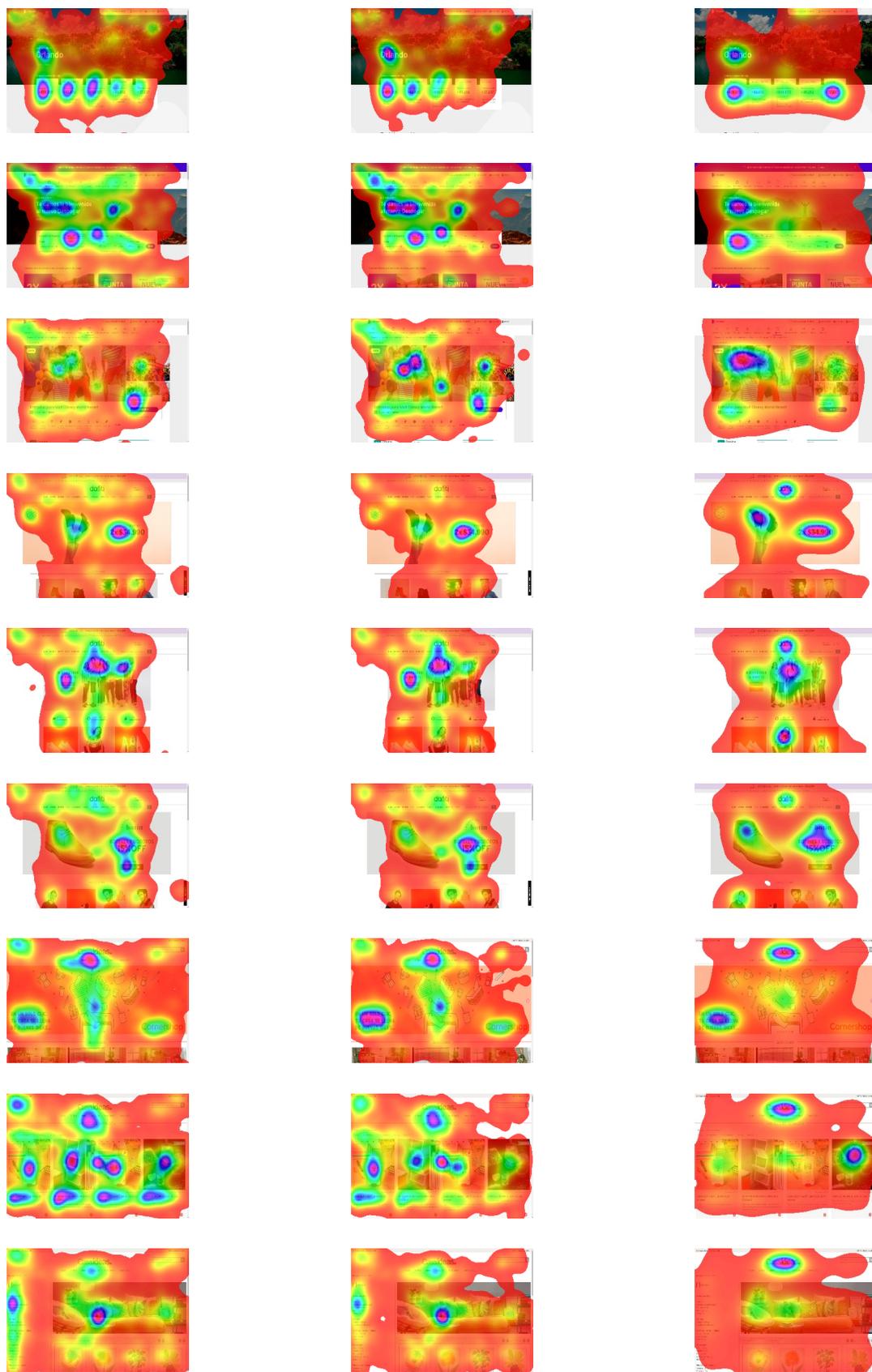


Figura 5.18: Por fila, Ground-Truth 10 y 5 segundos de visualización y saliencia entregada por SAM-Resnet respectivamente para una imagen en particular Conjunto 1

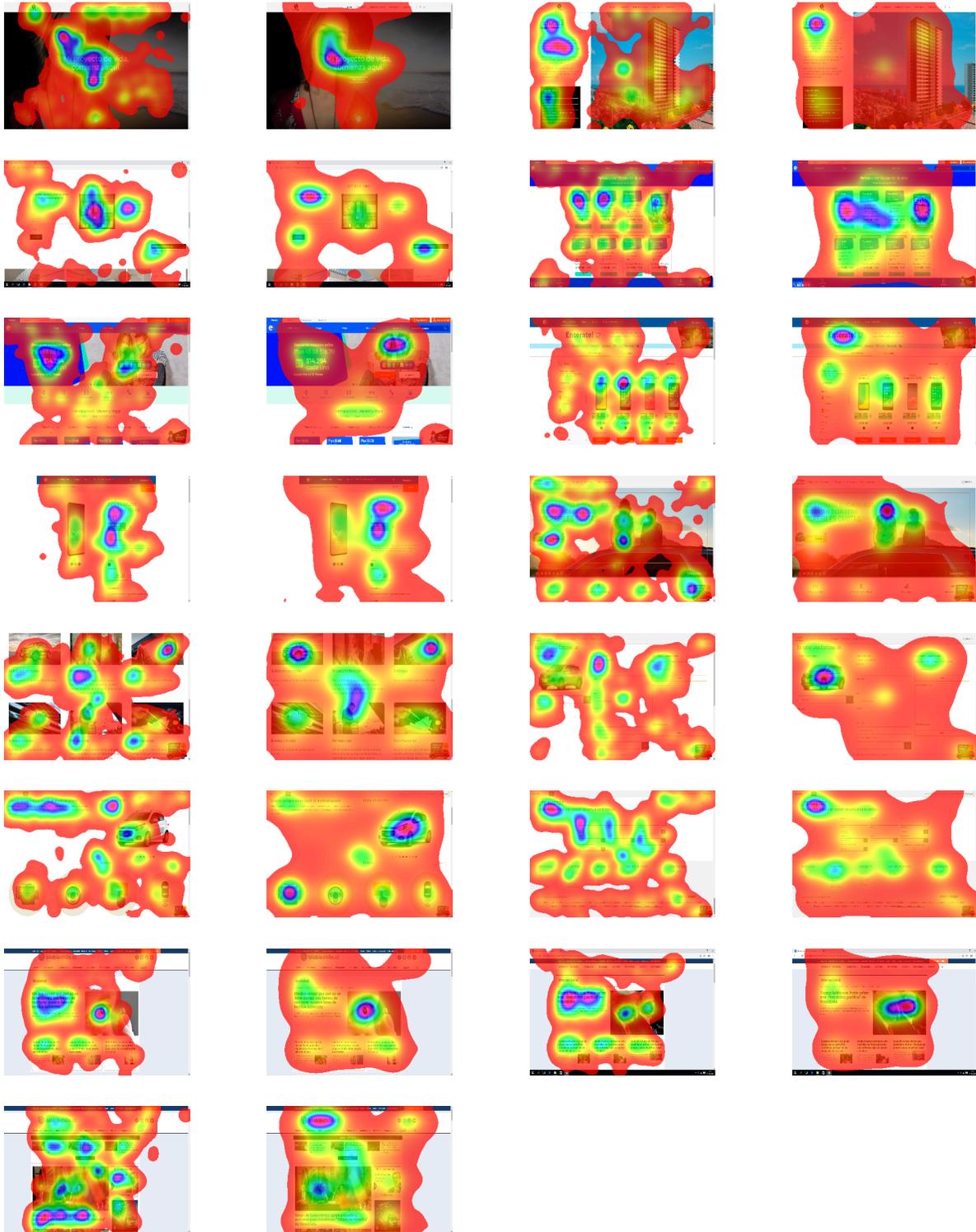


Figura 5.19: Por fila, dos pares de imágenes Ground Truth vs SAM-Resnet respectivamente, Conjunto 2 de imágenes

CAPITULO 6: CONCLUSIONES

6.1. Conclusión Investigación

Si bien numéricamente los datos muestran similitudes en los segmentos 1 y 2 y diferencias significativas de estos dos con el segmento 3, para la mayoría de los enfoques realizados en los análisis estadísticos, no se encuentra un patrón fuertemente definido para lograr caracterizar a cada uno de estos segmentos.

Lo anterior pues, si bien la entropía indica que las personas más adultas tienen una menor exploratividad (menor entropía), en términos de mapas de saliencia promedio no se ve reflejado este patrón con claridad. En algunos casos existen varias zonas de mayor probabilidad de saliencia de este segmento y en otros casos no, por lo que hace difícil la generalización de esta característica a las personas pertenecientes al segmento 3.

Para cada segmento, existe un alto acuerdo Intra-Grupo en las zonas de fijaciones ($AUC > 0.83$ para los 3 casos), pero no es posible concluir cual de los 3 segmentos tiene un mayor acuerdo estadísticamente, por lo que tampoco es posible declarar una conclusión fuerte de esto ni generalizar un patrón. Por otro lado, el acuerdo Inter - Segmentos de edad muestra diferencias significativas en las zonas de fijación entre los segmentos S1-S3 y S2-S3 siendo consistente con los resultados hallados en la correlación de Pearson de los mapas de saliencia, en donde el segmento 1 con el 2 tienen una alta correlación (r promedio 0.92), mientras que los segmentos S1-S3 y S2-S3 tienen valores más bajos ($r < 0.81$ en ambos casos) y no es posible rechazar la hipótesis de que las correlaciones de estas dos últimas combinaciones (S2-S3, S1-S3) son iguales ($p - valor > 0.25$).

Las diferencias de los dos primeros segmentos con el 3 pueden deberse a la alta variabilidad que tiene este último segmento en el comportamiento visual de los individuos que lo componen, a diferencia de las varianzas del segmento 1 con 2, esto se puede ver en la columna *std* en Tabla 5.12, en las combinaciones que aparece el segmento 3 la desviación estándar es más alta que en los otros casos.

En cuanto al sesgo central, los resultados son similares a los hallados en las métricas anteriores, en donde el segmento 3 logra ser estadísticamente diferente del 1 y 2, teniendo un menor sesgo central (mayor distancia euclídeana).

Finalmente, las diferencias halladas discutidas en los párrafos anteriores, para cada una de las métricas estudiadas, sólo son válidas para un subconjunto de páginas web pequeño

del conjunto de páginas total mostrado en el experimento, por lo que las conclusiones no son generalizables para la mayoría de páginas webs. Esto sugiere que para a futuro, el enfoque de estudio en la segmentación por edad de la visualización en páginas web debe considerar otros patrones u otro enfoque que también sea cuantificable y generalizable a un modelo matemático.

En cuanto a la evaluación del modelo predictor usado actualmente por el proyecto Akori, en términos generales cualitativos, se observa un patrón en donde los sujetos tienden a mirar los precios y texto de los artículos por sobre el artículo en sí, siendo lo más llamativo dentro de este tipo de páginas. El modelo SAM-Resnet tiende a hacer lo contrario, sin embargo, existen casos en los cuales si acierta bien, por lo que no es un error sistemático del modelo. El modelo podría ser reentrenado con esta data del experimento para ver si logra aprender el patrón mencionado en las líneas anteriores.

Por otro lado, en términos generales, SAM-Resnet tiene un buen desempeño en pronosticar las zonas que más llaman la atención en las páginas web de eCommerce para los primeros cinco y diez segundos de visualización, logrando promedios de *AUC-Judd* de 0.75 y 0.78 con desviaciones estándar de 0.05, siendo levemente más preciso para los cinco primeros segundos. Se rechaza una de las hipótesis del trabajo, en donde se creía que la precisión del modelo estaría muy por debajo de la que se declara en el *MIT saliency benchmark* (0.88 y 0.04 para el promedio y *std* respectivamente), debido a que el modelo esta pre entrenado con imágenes reales y no con páginas webs, creyéndose que el comportamiento visual en este último caso sería distinto.

Es probable la presencia de un sesgo en el experimento realizado, pues, el procedimiento, no captura la intención del usuario al momento de estar observando las páginas webs mostradas, habiendo un tiempo determinado en donde se está obligado a observar cada página por dicho lapso. Se cree que la saliencia en sí puede estar influenciada por el comportamiento del usuario en la web, por lo que los modelos de saliencia para web deben incluir esta componente para lograr decir algo más allá de cada saliencia que entregan, y así, ser más precisos al momento de entregar un servicio por parte del proyecto Akori.

El enfoque para las saliencias de páginas web no debe incluir solamente los elementos que componen estas últimas, pues, como se comprobó en este estudio, el modelo actual del proyecto Akori es lo suficiente robusto para lograr predecir con precisión las zonas que más llaman la atención en páginas de eCommerce, por lo que es necesario capturar ahora la intención del usuario.

Uno de los valores de este trabajo de investigación es el descubrimiento de la gran variabilidad en el comportamiento visual de personas chilenas en webs de eCommerce independiente de su rango etario, en donde es difícil encontrar un patrón en la saliencia que sea diferente por edad y a la vez cuantificable. En la muestra estudiada, para cada rango de edad existen personas que poseen altos, medios y bajos niveles de inspección en la visualización de las páginas web. Esto puede deberse al efecto de la publicidad en las personas en donde esta última no necesariamente impacta de la misma forma a todo tipo de público siendo más interesante para algunos y menos para otros.

Por otra parte, se realizó una validación del modelo SAM-Resnet en páginas de eCommer-

ce. Este modelo fue entrenado en parte con data de seguimiento ocular de 120 personas (40 hombres, 80 mujeres, diferentes etnias, promedio de edad 20 años, $std= 1.65$) en imágenes de diferentes contextos [74] [14], en donde no se incluyen páginas web. Se comprobó que el modelo es lo suficientemente bueno para ser usado en pronosticar saliencias en páginas web de eCommerce con usuarios chilenos de distinto rango etario (promedio 29 años, $std= 10.5$). En caso de que este modelo logre privilegiar precios y textos en la saliencia, mejoraría considerablemente la precisión de lo que realmente ven los usuarios.

En este Trabajo de Título hubo un levantamiento de información acerca del uso del dispositivo de *Eye Tracker* con un software que es distinto al original (*Captive-NeuroLab*) por lo tanto no se sabía como usarlo, ni la interpretación de los datos crudos exportados. En el manual de usuario no se especifica como por lo que se descubrió a prueba y error. Se conocieron ciertos percances del instrumento para servir a modo de retroalimentación cuando se vuelva a utilizar. También se crearon códigos útiles para el momento del pre tratamiento y limpieza de la data proveniente del *Eye Tracker*, así como también la creación de los mapas de saliencia, calor y el cálculo de métricas.

Finalmente, es importante destacar la representatividad de los descubrimientos hechos en este estudio debido a que el tamaño de la muestra es grande (80 personas) y se asemeja bastante a las investigaciones de la línea realizadas en el mundo en donde los tamaños oscilan entre 70 – 120 personas. Obtener conjuntos de datos de estos tamaños es costoso y requiere una cantidad de tiempo considerable y buena voluntad de las personas a participar, pues como fue el caso de esta investigación, los sujetos no recibieron un pago al realizar el experimento, lo que dificultó aún más la captación de estos. El conjunto de data proveniente de *Eye Tracking* de las 80 personas junto a los códigos respectivos de las métricas y limpieza de los datos queda disponible en el Web Intelligence Centre para futuros trabajos en el proyecto Akori.

6.2. Potencial Comercial

Aunque los instrumentos de *Eye Tracking* logran una alta precisión de predicción, no siempre son una opción disponible. Por lo tanto, el uso de modelos computacionales ha ganado importancia en las últimas décadas.

Es importante saber bien qué tipo imágenes y objetos llaman la atención para efectos de saliencias en páginas web, esto es debido a que las ilustraciones deben ser alrededor de 50% informativas, de tal manera que el usuario sólo con ver imágenes, sea capaz de crearse una idea de lo que esta viendo y donde se encuentra lo que el busca. Esto hace más rápida la navegación y más intuitiva, pues es necesario que en el menor tiempo posible, el usuario sepa bien donde está parado y hacia donde ir para obtener lo que quiere. Esta última información fue corroborada por Crístian Ampuero, diseñador web del Web intelligence Centre.

Los sitios de eCommerce tienen muchas dimensiones desde el punto de vista del usuario final. Hacer que los sitios web sean más usables y mejorar la experiencia de los usuarios representa un paso importante cuando se desea evitar que se vayan del sitio. Algunos autores [75] consideran que las características principales de las nuevas tecnologías son la novedad,

la capacidad de adaptarse y remodelar continuamente según las necesidades de los usuarios y los descubrimientos tecnológicos en el campo. Una nueva tendencia [76] también incluye tecnologías de TI verdes y sistemas de información verdes para una empresa adecuada.

Lo anterior sugiere que es importante estar constantemente actualizado en saber la forma en que los usuarios navegan en las páginas web y cuáles son las cosas que más le llaman la atención dentro de ellas. Estas cosas no necesariamente son las mismas para todas las personas, por lo que tener alguna variable diferenciadora, como por ejemplo la edad, sirve como insumo para el remodelamiento de las páginas web. El presente estudio sirve como un primer acercamiento en mejorar una herramienta que sirve en el remodelamiento de una web (proyecto Akori).

Por otro lado, existe una nueva tendencia la cual mezcla las redes sociales con el eCommerce. Este concepto es llamado s-Commerce y su definición es la siguiente [77]: un “comercio electrónico basado en múltiples usuarios” [78] que proporciona una experiencia de compra mejorada y habilitada por la tecnología [79] y presenta una variedad de actividades comerciales y sociales [80]. Como tales, las características clave asociadas con el concepto son (1) la opción de vender a través de plataformas de redes sociales, (2) agregar características a los sitios web de comercio electrónico (eCommerce) existentes que permiten la interacción del usuario (como calificar productos o dejar comentarios) y/o (3) compartir contenido con personas que tienen un interés similar [80] [79] [81].

El concepto de s-Commerce se puede describir como una subcategoría de eCommerce, con “características interactivas sociales” (por ejemplo, dar “me gusta”, compartir, comentar y re-twittear). Si bien a las redes sociales les resulta difícil llevar transacciones comerciales a sus plataformas [82], el desarrollo de los “nuevos modelos de negocios basados en tecnología” [82] están resultando difíciles de implementar y comprender por los minoristas de eCommerce.

El modelo de negocio actual para el eCommerce, que se centra en el diseño de navegación y las tecnologías de la Web 1.0 para respaldar las transacciones en línea [83], no se centra en los nuevos tipos de interacción que facilitan las herramientas de la Web 2.0 [82] por lo tanto, si las plataformas de eCommerce incluyen características sociales en sus interfaces, es fundamental que consideren las interacciones sociales en línea [82] y las características de diseño de s-Commerce.

De acuerdo con lo anterior esta área requiere un constante estudio y herramientas que sirvan de apoyo en la construcción de sitios web óptimos e integrales basados en las preferencias y comodidad de los usuarios, por lo que el proyecto Akori posee un potencial a futuro como una empresa capaz de entregar un servicio de herramienta para diseñadores webs.

6.3. Recomendaciones y Trabajo Futuro

Debido a la metodología empleada en este estudio, no es posible caracterizar la saliencia por tipos de elementos dentro de cada página web, por ejemplo, caras, objetos, colores, entre otros. Es necesario, como trabajo futuro, saber que elemento es lo primero que observa el

usuario para una página web en particular, sujeto a su intención al ser expuesto a una web. Luego de esto, identificar las características de estos primeros elementos segmentados por variables demográficas como sexo o edad. En caso de este último, utilizar segmentos de edad alejados (por ejemplo 18-24 y 50-65). La idea para el futuro es crear un sistema capaz de hacer este reconocimiento de características de manera automático.

Con lo anterior en mano, Akori sería capaz de entregar una especie de informe a modo de servicio, en donde se detalle todas estas características llamativas en los objetos salientes de la página web del cliente. Esto sería de utilidad a los diseñadores web, pues, podrían tomar estas características y ponerlas dentro de objetos de acción y así rediseñar la página para hacerla más efectiva.

Una posible metodología para realizar lo anterior es la siguiente:

Calcular el mapa de saliencia total de todas las personas para una determinada imagen. Luego, seleccionar las zonas mas salientes, es decir, las que tengan más alto nivel de probabilidad de atención. Con lo anterior, es posible etiquetar manualmente estas zonas más salientes para que en un paso posterior, se contabilicen las fijaciones de cada persona de manera individual en estas zonas etiquetadas (Ver Figura 6.1). De esta forma, en caso de que las zonas de mayor saliencia sean precios y texto, se podría comparar distintos grupos de personas para ver qué grupo se fijan más en qué cosa.

Con lo anterior, el siguiente paso sería encontrar la forma de que un modelo de saliencia logre capturar cada una de las características anteriores, para posteriormente ser calibrado en función de las preferencias de cada grupo de personas estudiado. Así, se lograría un modelo clasificador de saliencia por grupos específicos.

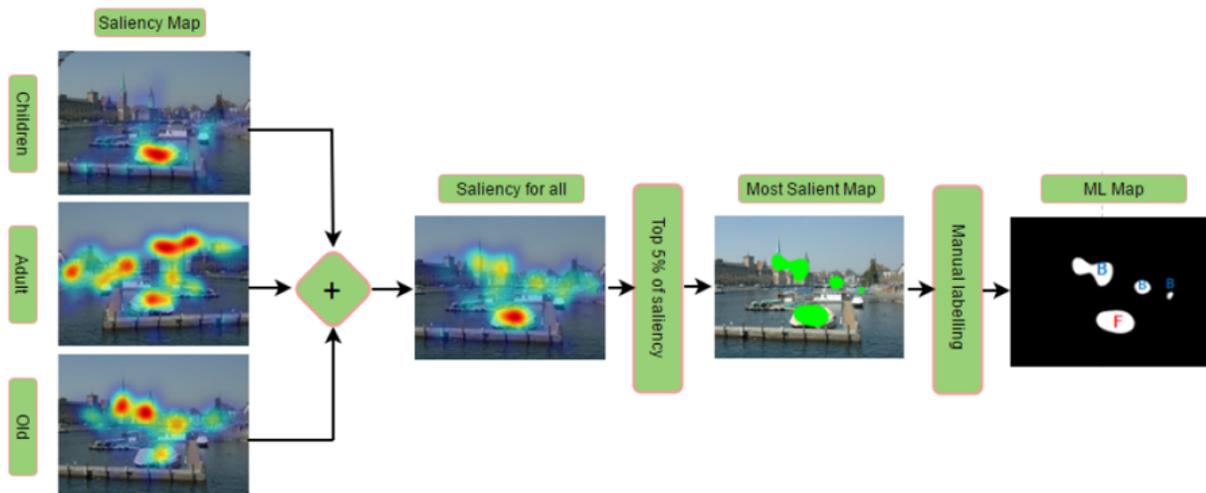


Figura 6.1: Ejemplo etiqueta en saliencia. Fuente: Adaptado de [2]

Finalmente se deja propuesto el reentrenamiento del modelo SAM-Resnet con la data obtenida del experimento realizado en este estudio. Con esto se puede evaluar si su desempeño aumenta y si es que logra capturar cuando privilegiar texto o precios por sobre imágenes de objetos, en los casos en que estos dos últimos se encuentren presentes al mismo tiempo en

una página web.

GLOSARIO

Web: Forma coloquial de referirse a la “www” que viene de “World Wide Web”. No es un sinónimo de internet, la web es un subconjunto del internet, que es la verdadera red de redes donde toda la información reside. Consiste en el conjunto de todas las páginas web que pueden ser accedidas usando un buscador web. El *HTTP* (*Hyper Text Transfer Protocol*) es el método utilizado para transferir páginas web al computador [84].

Web 1.0: La web del siglo 20. Se caracterizaba principalmente por ser unidireccional y realizada sobre contenidos estáticos. Las primeras páginas publicaban contenidos de texto que, una vez publicados, no se actualizaban salvo que el “webmaster” modificase dichos contenidos y volviese a subir la web de nuevo a internet [85].

Web 2.0: Se asentó a mediados de la primera década del siglo 21. Sustentada bajo unas conexiones a internet evolucionadas y mejores herramientas para desarrollar web, mejores servidores, entre otros. La web 2.0, también denominada “la red social”, llena Internet de blogs, wikis, foros y finalmente, redes sociales. El objetivo de la web 2.0 es el compartimiento del conocimiento, es la web colaborativa y ha sido uno de los atractivos principales para atraer a usuarios [85].

Web 3.0: Es la web semántica, la web de la nube, la web de las aplicaciones y la web multidispositivo. Hoy en día ya no solamente se utilizan ordenadores para conectarse a Internet. Los tablets, los smartphones, e incluso los mapas interactivos, algunas tiendas, y en un futuro no lejano la automoción estará consumiendo Internet. La web 3.0 se presenta como una web inteligente y principalmente aprovecha la nube para prestar servicios al usuario y eliminar su necesidad de disponer de sistemas operativos complejos y grandes discos duros para almacenar su información [85].

Páginas web: Las páginas web son las que constituyen la web. Estos documentos están escritos en *HTML* (*Hyper Text Markup Language*) y son traducidos por el navegador web. Pueden ser estáticas, es decir, muestran el mismo contenido cada vez que son accedidas, o dinámicas, cuando cuyo contenido cambia. Una página web no es lo mismo que un sitio web, un sitio web es una colección de páginas web [86].

Archivo log o registro: Un *log* es un archivo de texto en el que constan cronológicamente los acontecimientos que han ido afectando a un sistema informático (programa, aplicación, servidor, entre otros), así como el conjunto de cambios que estos han generado [87]. Existen diferentes tipos de archivos de *log* dependiendo de la información. Por ejemplo, existe un archivo de *log* del sistema, un archivo de *log* para los mensajes de seguridad y un archivo de

log para las tareas cron [88].

Atención Visual: Se entiende como un conjunto de áreas neuronales que ejecutan distintas operaciones para procesar la información captada por el ojo, considerando el contexto del usuario y su historia.

Fijación o fijación visual: Es el mantenimiento de la mirada visual en una única ubicación. Un animal puede exhibir una fijación visual si posee una fovea en la anatomía de su ojo. La fovea suele estar ubicada en el centro de la retina y es el punto de visión más claro.

Persecución suave de los movimientos oculares: Permiten que los ojos sigan de cerca un objeto en movimiento. Los movimientos oculares son lentos, continuos y se conjugan y se producen cuando un objeto fijo se mueve lentamente.

Sacada: Es un movimiento rápido y simultáneo de ambos ojos entre dos o más fases de fijación en la misma dirección. En contraste, en movimientos de búsqueda suaves, los ojos se mueven suavemente en lugar de en saltos. El fenómeno se puede asociar con un cambio en la frecuencia de una señal emitida o un movimiento de una parte del cuerpo o dispositivo.

Mirada (Gaze): La mirada, un movimiento coordinado de los ojos y la cabeza, se ha utilizado a menudo como un indicador de la atención en el comportamiento natural. Por ejemplo, un humano o un robot tiene que interactuar con los objetos circundantes y controlar la mirada para realizar una tarea mientras se mueve en el entorno. En este sentido, el control de la mirada compromete la visión, la acción y la atención simultáneamente para realizar la coordinación sensoriomotora necesaria para el comportamiento requerido (por ejemplo, alcanzar y agarrar) [89].

Ground Truth: La verdad fundamental es un término utilizado en varios campos para referirse a la información proporcionada por la observación directa (es decir, evidencia empírica) en oposición a la información proporcionada por la inferencia [90].

Saliencia visual: Cualidad estética de la forma de un objeto o una configuración que destaca uno o más de sus rasgos por sobre los demás y que, en consecuencia, llama la atención. Se puede decir que esta ocurre en la interacción entre sujeto y objeto, entre un agente y su medio ambiente porque si bien la saliencia visual es una propiedad del objeto, su percepción está a cargo del sujeto (agente). El sujeto (agente) es o no sensible a esa saliencia. Y esa sensibilidad puede ser biológica o cultural [91] [92].

Mapa de calor: Un mapa de calor es una representación gráfica de información en que los valores individuales de una matriz son mostrados mediante un código de colores. Si se aplica a páginas web, zonas concretas de esta resaltan en base a criterios como el número de clicks, o las áreas por las que se pasa con mayor frecuencia el puntero.

Generalmente, en colores cálidos como rojo y amarillo se destacan las zonas con mayor actividad y en colores fríos como azul y verde aquellos con menor interacción [93]. Para efectos de este estudio, el código de colores funciona de manera inversa a lo dicho en la línea anterior.

El objetivo de los mapas de calor es obtener del comportamiento de los navegantes en una

web y obtener datos útiles para mejorar aspectos de la misma: comprobar la visibilidad de ciertos elementos, o evaluar los mejores sitios para poner publicidad. Se necesita una muestra amplia de datos para poder hallar patrones de visionado y desplazamiento. El concepto del mapa de calor deriva de la técnica de *Eye Tracking* usada en los estudios de psicología y marketing para detectar dónde se posa la mirada al ver un texto o imagen (ver Figura 5.2).

Bottom-up: Es un enfoque para generar los mapas de saliencia que consiste en considerar como variables sólo la posible estimulación visual que puede generar una imagen determinada en un usuario, ya sea por los colores, orientación, distribución espacial y factores externos al usuario. Resultando finalmente un mapa influenciado exclusivamente por las características propias de la imagen y no por el usuario en particular [91].

Top-Down: Enfoque para determinar los mapas de saliencia. Este considera variables internas del sujeto expuesto a las imágenes, además de la imagen en sí. Este enfoque da cabida al análisis "*Task Driven*" donde se toma en cuenta el objetivo que tiene el usuario al momento de ser expuesto a una imagen, cuyos objetivos afectan el resultado del mapa dependiendo si el usuario al ser expuesto tiene como objetivo explorarla libremente, buscar alguna figura o color particular, entre otros. Finalmente se destaca que la experiencia previa del usuario al realizar dicho objetivo afecta su desempeño para este y por ende el resultado del mapa de saliencia [91].

Sesgo de profundidad: El sistema visual humano tiene la tendencia de centrarse antes en los objetos colocados en primer plano que los objetos en el fondo de la escena [94]. Gautier y Le Meur [95] investigaron la influencia de la disparidad en la prominencia en condiciones 3D. Sus resultados indican que las características de primer plano desempeñan un papel importante en el despliegue de la atención. Sin embargo, esta tendencia del sistema visual humano se explora a fondo técnicamente para los medios tridimensionales, sólo hay un estudio que investiga el sesgo de profundidad para los medios bidimensionales [2], por lo que no hay mayor información acerca de este patrón.

Sesgo de centralidad: La polarización central es la tendencia a desplazarse alrededor de las ubicaciones centrales de una escena mientras se ve la escena. Este es uno de los sesgos más fuertes reportados en muchos estudios de seguimiento ocular [96] [97]. La precisión de la predicción de muchos modelos computacionales de atención visual [98] [97] mejoraron cuando se incluyó el sesgo del centro en su marco de predicción. Por ejemplo, una burbuja gaussiana centrada en la mitad de la imagen, mejoró considerablemente el rendimiento de predicción del modelo de saliencia basado en el aprendizaje [98].

Redes Neuronales Convolucionales (CNN): Las redes neuronales convolucionales, o CNN, convolucionan las características aprendidas junto con la data y usan capas 2-dimensionales, haciendo la arquitectura adecuada para procesar data en 2D como lo son las imágenes. Las CNN son similares a las redes neuronales multicanal. Su principal ventaja es que cada parte de la red se le entrena para realizar una tarea, esto reduce significativamente el número de capas ocultas, por lo que el entrenamiento es más rápido. Además, presenta invarianza a la traslación de los patrones a identificar. En la convolución se realizan operaciones de productos y sumas entre la capa de partida y los "*n*" filtros (o kernel) que genera un mapa de características. Las características extraídas corresponden a cada posible ubicación del filtro en la imagen original [99].

Dispositivo de Eye Tracker: Es la herramienta con la que se realiza la acción de extraer información del usuario analizando sus movimientos oculares. Consiste en un aparato especial que lanza rayos infrarrojos a los ojos del que está mirando la imagen. La dirección que siguen estos rayos va de la pupila del usuario al aparato, permitiendo así calcular con precisión en que posición de la pantalla está mirando en el tiempo [100].

Tasa de conversión o *Conversion Rate* :La tasa de conversión es el porcentaje de usuarios que realizan una acción deseada. El ejemplo arquetípico de tasa de conversión es el porcentaje de visitantes del sitio web que compran algo en el sitio. Ejemplo: un sitio de comercio electrónico es visitado por 100,000 personas durante el mes de abril. Durante ese mes, 2,000 usuarios compraron algo del sitio. Por lo tanto, la tasa de conversión del sitio es $2,000 / 100,000 = 2\%$ [101].

Distribución Normal: Es una distribución con forma de campana simétrica. La media, moda y mediana coinciden, y esta distribución es descrita completamente por sus dos parámetros μ y σ . Las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de observaciones de los datos. En el intervalo $[\mu - \sigma, \mu + \sigma]$ se encuentra alrededor del 68.26% de los datos y en el intervalo $[\mu - 2\sigma, \mu + 2\sigma]$ el 95,44% . Una gran cantidad de fenómenos o variables biológicas, psicológicas o sociales, tienen una distribución Normal [102].

Prueba no paramétrica:Una prueba no paramétrica es una prueba de hipótesis que no requiere que la distribución de la población sea caracterizada por ciertos parámetros. Por ejemplo, muchas pruebas de hipótesis parten del supuesto de que la población sigue una distribución normal con los parámetros μ y σ (media y desviación respectivamente). Las pruebas no paramétricas no parten de este supuesto, de modo que son útiles cuando los datos son considerablemente no normales y resistentes a transformaciones [103].

Bibliografía

- [1] “Tobii Pro X2-30 screen-based eye tracker,” Jun. 2015. [En línea]. Disponible en: <https://www.tobii.com/product-listing/tobii-pro-x2-30/>
- [2] O. Krishna and K. Aizawa, “Age-adapted Saliency Model with Depth Bias,” in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '17. New York, NY, USA: ACM, 2017, pp. 5:1–5:8. [En línea]. Disponible en: <http://doi.acm.org/10.1145/3119881.3119885>
- [3] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, “Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model,” *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5142–5154, 2018.
- [4] G. Slanzi, J. A. Balazs, and J. D. Velásquez, “Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention,” *Information Fusion*, vol. 35, pp. 51–57, May 2017. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1566253516300756>
- [5] J. Jadue, G. Slanzi, L. Salas, and J. D. Velásquez, “Web User Click Intention Prediction by Using Pupil Dilation Analysis,” in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 1, Dec. 2015, pp. 433–436.
- [6] G. Slanzi, J. Balazs, and J. D. Velásquez, “Predicting Web User Click Intention Using Pupil Dilation and Electroencephalogram Analysis,” in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Oct. 2016, pp. 417–420.
- [7] L. E. Dujovne and J. D. Velásquez, “Design and Implementation of a Methodology for Identifying Website Keyobjects,” in *Knowledge-Based and Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, J. D. Velásquez, S. A. Ríos, R. J. Howlett, and L. C. Jain, Eds. Springer Berlin Heidelberg, 2009, pp. 301–308.
- [8] J. D. Velásquez, L. E. Dujovne, and G. L’Huillier, “Extracting significant Website Key Objects: A Semantic Web mining approach,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 8, pp. 1532–1541, Dec. 2011. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S095219761100025X>

- [9] J. D. V. Silva, “Improvement of a Methodology for Website Keyobject Identification through the Application of Eye-Tracking Technologies,” in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 3, Dec. 2012, pp. 59–63.
- [10] P. Loyola, G. Martinez, K. Muñoz, J. D. Velásquez, P. Maldonado, and A. Couve, “Combining eye tracking and pupillary dilation analysis to identify Website Key Objects,” *Neurocomputing*, vol. 168, pp. 179–189, Nov. 2015. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0925231215008061>
- [11] G. Slanzi, C. Aracena, and J. D. Velásquez, “Eye Tracking and EEG Features for Salient Web Object Identification,” in *BIH*, 2015.
- [12] H. Yasuda, S. A. Ríos, A. Bassi, J. D. Velásquez, and T. Aoki, “Towards the identification of keywords in the web site text content: A methodological approach,” *International Journal of Web Information Systems*, vol. 1, no. 1, pp. 53–57, Feb. 2005. [En línea]. Disponible en: <https://www.emeraldinsight.com/doi/abs/10.1108/17440080580000083>
- [13] “Akori.” [En línea]. Disponible en: <https://www.akoriproject.cl/>
- [14] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, “Mit saliency benchmark,” <http://saliency.mit.edu/>.
- [15] E. Erdem and A. Erdem, “Visual saliency estimation by nonlinearly integrating features using region covariances,” *Journal of Vision*, vol. 13, no. 4, pp. 11–11, Mar. 2013. [En línea]. Disponible en: <https://jov.arvojournals.org/article.aspx?articleid=2121371>
- [16] J. Harel, C. Koch, and P. Perona, “Graph-Based Visual Saliency,” *Adv. Neural Inform. Process. Syst.*, p. 8, 2006.
- [17] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [19] R. D. Wright, *Visual Attention*. Oxford University Press, 1998, google-Books-ID: HktnDAAAQBAJ.
- [20] N. Sidorakis, G. A. Koulteris, and K. Mania, “Binocular eye-tracking for the control of a 3d immersive multimedia user interface,” in *2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*, Mar. 2015, pp. 15–18.
- [21] “Modelling the interaction of attention and emotion - IEEE Conference Publication.” [En línea]. Disponible en: <https://ieeexplore.ieee.org/abstract/document/1556129>

- [22] C. Balkenius, “Attention, Habituation and Conditioning: Toward a Computational Model,” *Cognitive Science Quarterly*, p. 29, 2000.
- [23] “Frontiers | Developmental Changes in Natural Viewing Behavior: Bottom-Up and Top-Down Differences between Children, Young Adults and Older Adults | Psychology.” [En línea]. Disponible en: <https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00207/full>
- [24] “The maturation of eye movement behavior: Scene viewing characteristics in children and adults - ScienceDirect.” [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0042698914001850>
- [25] “Development of eye-movement control - ScienceDirect.” [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0278262608002649>
- [26] O. Krishna, A. Helo, P. Rämä, and K. Aizawa, “Gaze distribution analysis and saliency prediction across age groups,” *PLOS ONE*, vol. 13, no. 2, p. e0193149, Feb. 2018. [En línea]. Disponible en: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193149>
- [27] “Development of visual perception - Johnson - 2011 - Wiley Interdisciplinary Reviews: Cognitive Science - Wiley Online Library.” [En línea]. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.128>
- [28] A. Chandna, “Natural history of the development of visual acuity in infants,” *Eye*, vol. 5, no. 1, pp. 20–26, Jan. 1991. [En línea]. Disponible en: <https://www.nature.com/articles/eye19914>
- [29] A. Roucoux, C. Culee, and M. Roucoux, “Development of fixation and pursuit eye movements in human infants,” *Behavioural Brain Research*, vol. 10, no. 1, pp. 133–139, Oct. 1983. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/0166432883901596>
- [30] B. Luna, K. Velanova, and C. F. Geier, “Development of eye-movement control,” *Brain and Cognition*, vol. 68, no. 3, pp. 293–308, Dec. 2008. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0278262608002649>
- [31] J. Fukushima, T. Hatta, and K. Fukushima, “Development of voluntary control of saccadic eye movements: I. Age-related changes in normal children,” *Brain and Development*, vol. 22, no. 3, pp. 173–180, May 2000. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0387760400001017>
- [32] E. L. Irving, M. J. Steinbach, L. Lillakas, R. J. Babu, and N. Hutchings, “Horizontal Saccade Dynamics across the Human Life Span,” *Investigative Ophthalmology & Visual Science*, vol. 47, no. 6, pp. 2478–2484, Jun. 2006. [En línea]. Disponible en: <https://iovs.arvojournals.org/article.aspx?articleid=2125279>
- [33] C. Klein and F. Foerster, “Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years,” *Psychophysiology*,

vol. 38, no. 2, pp. 179–189, Mar. 2001. [En línea]. Disponible en: <https://www.cambridge.org/core/journals/psychophysiology/article/development-of-prosaccade-and-antisaccade-task-performance-in-participants-aged-6-to-26-years/0742F7628C66473AC5A86C3213E72C2A>

- [34] J. Li and W. Gao, “Learning-Based Visual Saliency Computation,” in *Visual Saliency Computation: A Machine Learning Perspective*. Cham: Springer International Publishing, 2014, pp. 101–149. [En línea]. Disponible en: https://doi.org/10.1007/978-3-319-05642-5_5
- [35] T. Mei, X.-S. Hua, L. Yang, and S. Li, “VideoSense: Towards Effective Online Video Advertising,” in *Proceedings of the 15th ACM International Conference on Multimedia*, ser. MM ’07. New York, NY, USA: ACM, 2007, pp. 1075–1084, event-place: Augsburg, Germany. [En línea]. Disponible en: <http://doi.acm.org/10.1145/1291233.1291467>
- [36] C.-H. Chang, K.-Y. Hsieh, M.-C. Chiang, and J.-L. Wu, “Virtual spotlighted advertising for tennis videos,” *Journal of Visual Communication and Image Representation*, vol. 21, no. 7, pp. 595–612, Oct. 2010. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1047320310000386>
- [37] R. Espinosa, “¿Qué es el modelo AIDA en Marketing?” [En línea]. Disponible en: <https://robertoespinosa.es/2017/04/11/modelo-aida-marketing-metodo/>
- [38] M. Mancas, “Relative Influence of Bottom-Up and Top-Down Attention,” in *Attention in Cognitive Systems*, ser. Lecture Notes in Computer Science, L. Paletta and J. K. Tsotsos, Eds. Springer Berlin Heidelberg, 2009, pp. 212–226.
- [39] M. C. Chen, J. R. Anderson, and M. H. Sohn, “What Can a Mouse Cursor Tell Us More?: Correlation of Eye/Mouse Movements on Web Browsing,” in *CHI ’01 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’01. New York, NY, USA: ACM, 2001, pp. 281–282, event-place: Seattle, Washington. [En línea]. Disponible en: <http://doi.acm.org/10.1145/634067.634234>
- [40] “Feng-GUI - Effective Visuals.” [En línea]. Disponible en: <http://www.feng-gui.com/>
- [41] “VAS: 3m Visual Attention Software (VAS).” [En línea]. Disponible en: https://www.3m.com/3M/en_US/company-us/all-3m-products/~VAS-3M-Visual-Attention-Software-VAS-/?N=5002385+3293010368&rt=rud
- [42] W. L. GmbH (EyeQuant), “We make design measurable.” [En línea]. Disponible en: <https://www.eyequant.com>
- [43] “Ittention - La plateforme de test d’impact visuel.” [En línea]. Disponible en: <http://www.ittention.com/>
- [44] “EyePredict: Home.” [En línea]. Disponible en: <https://eye-predict.com/>
- [45] Wackerly, Dennis, W. M. III, and R. L. Scheaffer, *Estadística Matemática con Aplicaciones Séptima Edición*. Thomson/Brooks-Cole, 2008. [En línea]. Dispo-

nible en: https://www.academia.edu/37461181/_Wackerly_Mendenhall_Scheaffer_Estadistica_Matematica_con_Aplicaciones

- [46] “Tipos de pruebas t.” [En línea]. Disponible en: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/tests-of-means/types-of-t-tests/>
- [47] M. B. Brown and A. Forsythe, “Robust Tests for Equality of Variances,” *JASA. Journal of the American Statistical Association*, vol. 69, Sep. 1974.
- [48] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3-4, pp. 591–611, Dec. 1965. [En línea]. Disponible en: <https://academic.oup.com/biomet/article/52/3-4/591/336553>
- [49] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, Mar. 1947. [En línea]. Disponible en: <https://projecteuclid.org/euclid.aoms/1177730491>
- [50] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis,” *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, Dec. 1952. [En línea]. Disponible en: <https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441>
- [51] A. Dinno, “Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn’s Test,” *Stata Journal*, vol. 15, pp. 292–300, Jan. 2015.
- [52] O. J. Dunn, “Multiple Comparisons Using Rank Sums,” *Technometrics*, vol. 6, no. 3, pp. 241–252, Aug. 1964. [En línea]. Disponible en: <https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1964.10490181>
- [53] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556 [cs]*, Sep. 2014, arXiv: 1409.1556. [En línea]. Disponible en: <http://arxiv.org/abs/1409.1556>
- [54] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. [En línea]. Disponible en: http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- [55] “CAPTIV NeuroLAB – Your unmatched solution for human behavior analysis.” [En línea]. Disponible en: <http://captiv-neurolab.com/>
- [56] “Google Colaboratory.” [En línea]. Disponible en: <https://colab.research.google.com/notebooks/welcome.ipynb?hl=es-419>
- [57] K. McSpadden, “You Now Have a Shorter Attention Span Than a Goldfish,” *Time*, 2015. [En línea]. Disponible en: <https://time.com/3858309/attention-spans-goldfish/>

- [58] R. Weatherhead, “Say it quick, say it well – the attention span of a modern internet consumer,” *The Guardian*, Feb. 2014. [En línea]. Disponible en: <https://www.theguardian.com/media-network/media-network-blog/2012/mar/19/attention-span-internet-consumer>
- [59] R. Ruiz Moreno, “Diseño de un modelo de negocio para un servicio de análisis e indicadores sobre un contenido visual, asociado a proyecto AKORI,” Memoria de Título, Universidad de Chile, 2014. [En línea]. Disponible en: <http://repositorio.uchile.cl/handle/2250/130374>
- [60] O. Mir and M. Sofía, “Modelo de negocio para la comercialización de un servicio informático de análisis de imágenes, basado en Web Intelligence, exploración visual y neurociencia,” Memoria de Título, Universidad de Chile, 2018. [En línea]. Disponible en: <http://repositorio.uchile.cl/handle/2250/152508>
- [61] O. Le Meur and T. Baccino, “Methods for comparing scanpaths and saliency maps: strengths and weaknesses,” *Behavior Research Methods*, vol. 45, no. 1, pp. 251–266, Mar. 2013. [En línea]. Disponible en: <https://doi.org/10.3758/s13428-012-0226-9>
- [62] B. W. Tatler, “The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions,” *Journal of Vision*, vol. 7, no. 14, pp. 4–4, Nov. 2007. [En línea]. Disponible en: <https://jov.arvojournals.org/article.aspx?articleid=2122066>
- [63] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, “What Do Different Evaluation Metrics Tell Us About Saliency Models?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 3, pp. 740–757, Mar. 2019.
- [64] “Primer ranking de e-commerce revela quiénes lideran el negocio en Chile | La Tercera.” [En línea]. Disponible en: <https://www.latercera.com/noticia/e-commerce-chile/?fbclid=IwAR1LyZwMvuMkxY6UvvhR6BduckNyfHo4cYrxMjkbLsVRh7j9hUU-87LuM8w>
- [65] ecommerceday, “Se conocieron los Ganadores de los eCommerce Awards Chile 2018.” [En línea]. Disponible en: <http://www.ecommerceday.cl/2018/2018/2018/04/12/se-conocieron-los-ganadores-de-los-e-commerce-awards-chile-2018/>
- [66] C. Strello, “PRIMER RANKING E-COMMERCE EN CHILE,” Oct. 2017. [En línea]. Disponible en: https://www.scan.cl/2017/10/17/ranking-e-commerce-en-chile_scan-inteligencia-de-mercado-clientes-incognitos-estudios-de-mercado-inteligencia-competitiva-calidad-de-servicio-b2b/
- [67] “EyN: Dafiti apuesta por triplicar tamaño de su negocio en Chile en próximos tres años y potenciar ‘mobile.’” [En línea]. Disponible en: <http://www.economiaynegocios.cl/noticias/noticias.asp?id=528109&fbclid=IwAR2fsNndNIwGyRkRe06DppRM1VPw-qcYVNJWpSj3vjBwCOLXeSZbHRaD5pY>
- [68] “VIAJES FALABELLA, ES RECONOCIDA LÍDER EN ECOMMERCE,” Apr. 2018. [En línea]. Disponible en: <https://innovacionchilena.cl/viajes-falabella-es-reconocida->

- [69] “Chile - Supermercados Online,” Nov. 2017. [En línea]. Disponible en: <https://www.netrica.com/2017/11/10/chile-supermercados-online/>
- [70] “Groupon se convierte en una de las mayores empresas de e-commerce en América Latina.” [En línea]. Disponible en: <http://www.mundomaritimo.cl/noticias/groupon-se-convierte-en-una-de-las-mayores-empresas-de-e-commerce-en-america-latina>
- [71] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *2009 IEEE 12th International Conference on Computer Vision*, Sep. 2009, pp. 2106–2113.
- [72] J. M. Wolfe, “Guided Search 2.0 A revised model of visual search,” *Psychonomic Bulletin & Review*, vol. 1, no. 2, pp. 202–238, Jun. 1994. [En línea]. Disponible en: <https://doi.org/10.3758/BF03200774>
- [73] D. Estrategia, “Estudio GfK Adimark-Mercado Libre: 74% de los consumidores online chilenos aumentaron sus compras en el último año.” [En línea]. Disponible en: <http://www.estrategia.cl/texto-diario/mostrar/1040248/estudio-gfk-adimark-mercado-libre-74-consumidores-online-chilenos-aumentaron-compras-ultimo-ano>
- [74] A. Borji and L. Itti, “CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research,” *arXiv:1505.03581 [cs]*, May 2015, arXiv: 1505.03581. [En línea]. Disponible en: <http://arxiv.org/abs/1505.03581>
- [75] M. Georgescu and D. Popescul, “The Uncertainty of Using Web 2.0 Technologies in E-Government Development. Romania’s Case,” *Procedia Economics and Finance*, vol. 15, pp. 769–776, Jan. 2014. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S2212567114004857>
- [76] L.-D. Radu, “Green Information System for a Sustainable Enterprise,” *Sustainable Entrepreneurship and Investments in the Green Economy*, pp. 144–168, 2017. [En línea]. Disponible en: <https://www.igi-global.com/chapter/green-information-system-for-a-sustainable-enterprise/174472>
- [77] C. E. Henninger, N. Bürklin, and C. J. Parker, “Social Media’s Evolution in S-commerce,” in *Social Commerce: Consumer Behaviour in Online Environments*, R. Boardman, M. Blazquez, C. E. Henninger, and D. Ryding, Eds. Cham: Springer International Publishing, 2019, pp. 17–41. [En línea]. Disponible en: https://doi.org/10.1007/978-3-030-03617-1_2
- [78] T. Yamakami, “A view model of social commerce: The building blocks of next-generation e-commerce,” in *2014 14th International Symposium on Communications and Information Technologies (ISCIT)*, Sep. 2014, pp. 284–288.
- [79] J. Shen and L. B. Eder, “An Examination of Factors Associated with User Acceptance of Social Shopping Websites,” *International Journal of Technology and Human Interaction (IJTHI)*, vol. 7, no. 1, pp. 19–36, Jan. 2011. [En línea]. Disponible en: https://doi.org/10.1007/978-3-030-03617-1_2

//www.igi-global.com/article/examination-factors-associated-user-acceptance/49666

- [80] T.-P. Liang and E. Turban, “Introduction to the Special Issue Social Commerce: A Research Framework for Social Commerce,” *International Journal of Electronic Commerce*, vol. 16, no. 2, pp. 5–14, Dec. 2011. [En línea]. Disponible en: <https://doi.org/10.2753/JEC1086-4415160201>
- [81] D. Kim, “Under what conditions will social commerce business models survive?” *Electronic Commerce Research and Applications*, vol. 12, no. 2, pp. 69–77, Apr. 2013. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1567422312001238>
- [82] Y. Baghdadi, “A framework for social commerce design,” *Information Systems*, vol. 60, pp. 95–113, Aug. 2016. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0306437916301144>
- [83] H.-C. Ko, “Social desire or commercial desire? The factors driving social sharing and shopping intentions on social commerce platforms,” *Electronic Commerce Research and Applications*, vol. 28, pp. 1–15, Mar. 2018. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1567422317301047>
- [84] “WWW (World Wide Web) Definition.” [En línea]. Disponible en: <https://techterms.com/definition/www>
- [85] “Estudio Seijo.” [En línea]. Disponible en: <http://www.estudioseijo.com/noticias/web-10-web-20-y-web-30.htm>
- [86] “Web Page Definition.” [En línea]. Disponible en: <https://techterms.com/definition/webpage>
- [87] “¿Qué es un log?” [En línea]. Disponible en: <http://support.ankama.com/hc/es/articles/203790076--Qu%C3%A9-es-un-log->
- [88] “Archivos de registro.” [En línea]. Disponible en: <http://web.mit.edu/rhel-doc/3/rhel-sag-es-3/ch-logfiles.html>
- [89] M. Hayhoe and D. Ballard, “Eye movements in natural behavior,” *Trends in Cognitive Sciences*, vol. 9, no. 4, pp. 188–194, Apr. 2005. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1364661305000598>
- [90] “Ground truth,” Feb. 2019, page Version ID: 881421542. [En línea]. Disponible en: https://en.wikipedia.org/w/index.php?title=Ground_truth&oldid=881421542
- [91] A. Borji and L. Itti, “State-of-the-Art in Visual Attention Modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, Jan. 2013.
- [92] C. Koch and S. Ullman, “Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry,” in *Matters of Intelligence: Conceptual Structures*

- in Cognitive Neuroscience*, ser. Synthese Library, L. M. Vaina, Ed. Dordrecht: Springer Netherlands, 1987, pp. 115–141. [En línea]. Disponible en: https://doi.org/10.1007/978-94-009-3833-5_5
- [93] “¿Qué son los mapas de calor?” [En línea]. Disponible en: <http://www.ticbeat.com/tecnologias/que-son-los-mapas-de-calor/>
- [94] L. Jansen, S. Onat, and P. König, “Influence of disparity on fixation and saccades in free viewing of natural scenes,” *Journal of Vision*, vol. 9, no. 1, pp. 29–29, Jan. 2009. [En línea]. Disponible en: <https://jov.arvojournals.org/article.aspx?articleid=2122479>
- [95] J. Gautier and O. Le Meur, “A Time-Dependent Saliency Model Combining Center and Depth Biases for 2d and 3d Viewing Conditions,” *Cognitive Computation*, vol. 4, no. 2, pp. 141–156, Jun. 2012. [En línea]. Disponible en: <https://doi.org/10.1007/s12559-012-9138-3>
- [96] B. W. Tatler, “The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions,” *Journal of Vision*, vol. 7, no. 14, pp. 4–4, Nov. 2007. [En línea]. Disponible en: <https://jov.arvojournals.org/article.aspx?articleid=2122066>
- [97] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “SUN: A Bayesian framework for saliency using natural statistics,” *Journal of Vision*, vol. 8, no. 7, pp. 32–32, May 2008. [En línea]. Disponible en: <https://jov.arvojournals.org/article.aspx?articleid=2297284>
- [98] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *2009 IEEE 12th International Conference on Computer Vision*. Kyoto: IEEE, Sep. 2009, pp. 2106–2113. [En línea]. Disponible en: <http://ieeexplore.ieee.org/document/5459462/>
- [99] “Red neuronal Convolutacional CNN,” Jul. 2017. [En línea]. Disponible en: <http://www.diegocalvo.es/red-neuronal-convolutacional-cnn/>
- [100] E. Canelles, “¿Qué es el “Eye Tracking” y para qué nos sirve?” Jul. 2012. [En línea]. Disponible en: <https://www.solucionesc2.com/que-es-el-eye-tracking-y-para-que-nos-sirve/>
- [101] “Conversion Rate: Definition as used in UX and web analytics.” [En línea]. Disponible en: <https://www.nngroup.com/articles/conversion-rates/>
- [102] J. D. S, “LA DISTRIBUCIÓN NORMAL,” *Revista Chilena de Anestesia*, Dec. 2015. [En línea]. Disponible en: <http://revistachilenadeanestesia.cl/la-distribucion-normal/>
- [103] “Explicación de los métodos no paramétricos.” [En línea]. Disponible en: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/nonparametrics/supporting-topics/understanding-nonparametric-methods/>

Anexos

8.1. Datos demográficos participantes experimento

Edad	Cantidad	Hombre	Mujer
18	10	6	4
19	4	2	2
20	1	1	0
21	1	1	0
22	5	2	3
23	3	2	1
24	14	11	3
25	8	7	1
26	2	2	0
27	3	3	0
28	3	0	3
29	3	2	1
30	1	1	0
31	1	1	0
32	1	1	0
34	1	0	1
36	2	1	1
37	1	1	0
38	1	0	1
40	2	1	1
45	3	2	1
46	1	0	1
47	1	1	0
48	2	0	2
49	1	1	0
50	1	1	0
52	1	0	1
54	1	0	1
55	1	0	1
57	1	0	1
Total	80	50	30

Tabla 8.1: Datos demográficos participantes experimento

8.2. Consentimiento Informado

En las siguientes páginas se adjunta el consentimiento informado para la realización del experimento.

Comité de ética
Departamento de Ingeniería Industrial
Facultad de Ciencias Físicas y matemáticas
Universidad de Chile

Consentimiento Informado

Nombre de investigador principal: Dr. Juan Velásquez Silva
Institución: Web Intelligence Centre, DII. FCFM, Universidad de Chile
Teléfono: +56229784084

Se le entregará una copia del consentimiento informado completo.

Introducción

Mi nombre es Bastián Pardo Vergara, soy memorista de la carrera Ingeniería Civil Industrial y actualmente llevo a cabo la investigación a la cual usted ha sido invitado/a. A lo largo de este documento usted es libre de manifestar cualquiera de sus inquietudes respecto al procedimiento que se llevará a cabo, tanto hacia mí como con alguien con quien usted se sienta cómodo. Asimismo, puede tomarse el tiempo que requiera para reflexionar respecto a si desea participar del proyecto. Si no entiende alguna de las informaciones contenidas en el presente documento puede expresármelo en cualquier momento para explicarle la información en detalle.

Invitación a participar: Le estamos invitando a participar en un proyecto de investigación para la mejora de estructura y contenidos de sitios web.

Objetivos: Esta investigación tiene por objetivo estudiar la mirada de los usuarios al momento de observar páginas web a través de un experimento. El total de sujetos propuesto para realizar este estudio es de 40 a 60 personas.

Procedimientos: Si usted acepta participar en la investigación propuesta usted estará aceptando ser sometido/a, por una sola vez, al siguiente procedimiento: la medición del seguimiento ocular y dilatación pupilar mediante un sistema de cámara dispuesto en el monitor del computador (Eye Tracker Tobii X2-30). Este registro es no invasivo, y no produce daño ni efectos adversos. Durante todo el procedimiento no se le administrará ningún tipo de energía así como ningún tipo de fármaco. El experimento solo presenta una tarea, la cual es observar la pantalla mientras se le muestran sitios web. Usted es libre de retirarse de la tarea en cualquier momento a lo largo de esta, aunque no haya llegado a su fin. El investigador se encontrará en la misma sala que usted a lo largo del procedimiento, y usted puede solicitar a él en todo momento cualquier información o expresar cualquier necesidad que estime pertinente acerca del experimento.

En cuanto a datos demográficos, se almacenará el sexo, edad y ocupación del participante y en cuanto a la data proveniente del Eye Tracker se almacenará el seguimiento ocular y dilatación pupilar. Toda la data mencionada se guardará de manera anonimizada de tal manera que no sea posible identificar a los individuos mediante alguna técnica matemática. Adicionalmente, los registros extraídos del experimento podrían ser usados en el marco de esta investigación durante un plazo de dos años a partir de la fecha del presente documento.

Riesgos: Bajo el sistema de registro que utilizaremos no existen riesgos ni efectos adversos conocidos. El sistema de Eye Tracking es ampliamente utilizado en el mundo entero para fines de investigación, como en el caso de la tarea a la cual usted será sometido/a.

Costos: Las técnicas utilizadas en este proyecto no tienen costo alguno para usted.

Figura 8.1: Consentimiento informado pag 1

Beneficios: Los beneficios del presente estudio no irán en beneficio directamente de usted. El presente proyecto tiene por objetivo contribuir al conocimiento científico de como las miradas son conducidas por las personas.

Alternativas: La decisión de no participar del presente estudio no significará ningún prejuicio para su persona.

Compensación: No se considera la entrega de una compensación económica para usted en el presente estudio.

Confidencialidad: Toda la información derivada de su participación en este estudio será conservada en forma de estricta confidencialidad. Lo que incluye el acceso de los investigadores o agencias supervisoras de la investigación. Cualquier publicación o comunicación científica de los resultados de la investigación será completamente anónima.

Voluntariedad: Su participación en esta investigación es totalmente voluntaria y se puede retirar en cualquier momento comunicándolo al investigador.

Derechos del participante: Si usted requiere cualquier otra información sobre su participación en este estudio puede llamar a:

Memorista: Bastián Pardo Vergara, Teléfono +56997108145
Autoridad de la institución: Juan Velásquez Silva, Teléfono +56229784834

Conclusión: Después de haber recibido íntegramente y comprendido la totalidad de la información contenida en este documento, no teniendo actualmente ninguna duda respecto a la tarea a realizar ni las técnicas de medición, otorgo mi consentimiento para participar en el proyecto de investigación para la mejora de estructura y contenidos de sitios web.

_____	_____	_____
Nombre Sujeto	Firma	Fecha
Bastián Pardo Vergara		
_____	_____	_____
Nombre Memorista	Firma	Fecha

Figura 8.2: Consentimiento informado pag 2

8.3. Páginas Web Experimento

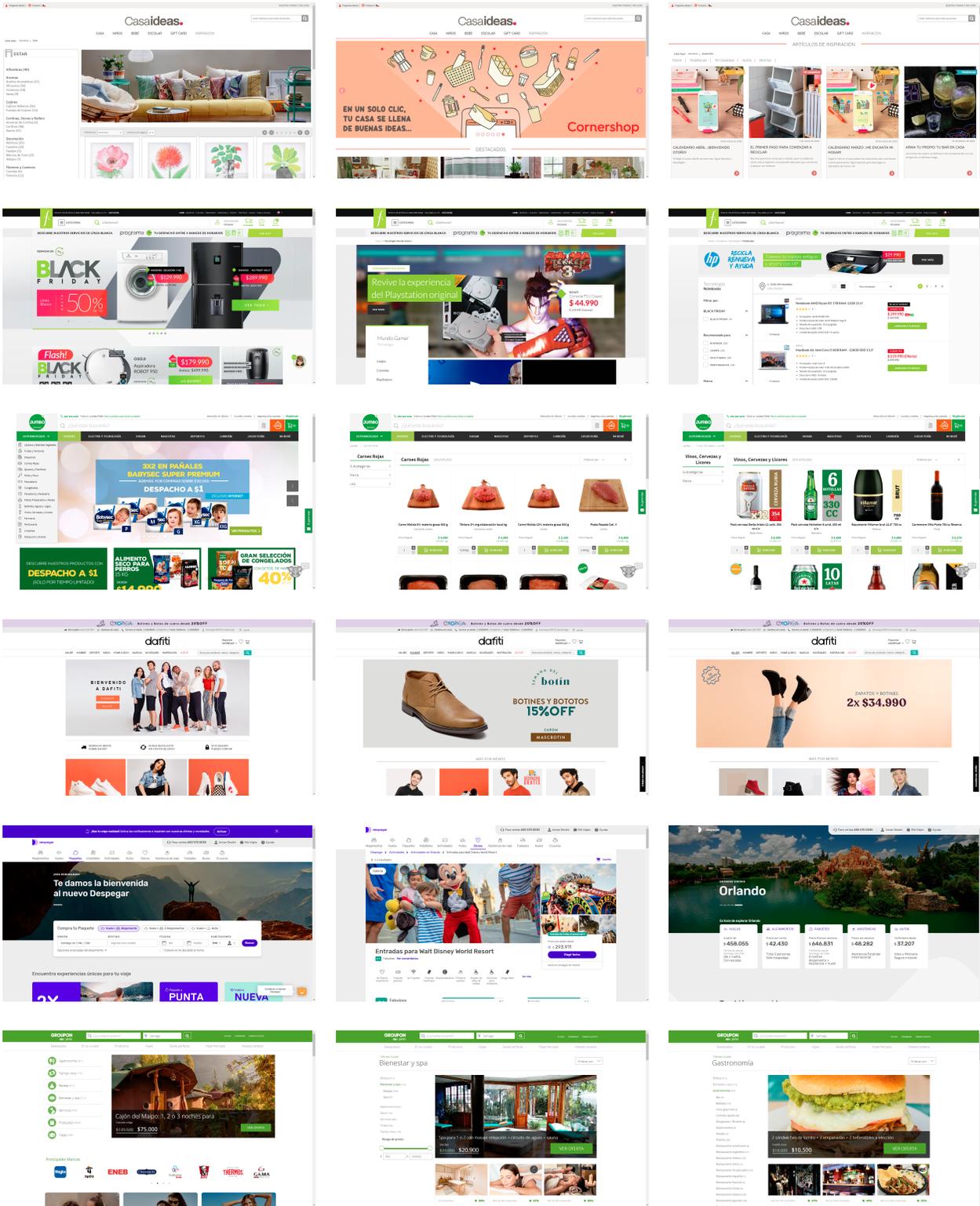


Figura 8.3: Conjunto de imágenes 1, web Casa & Ideas, Falabella, Jumbo, Dafiti, Despegar, Groupon

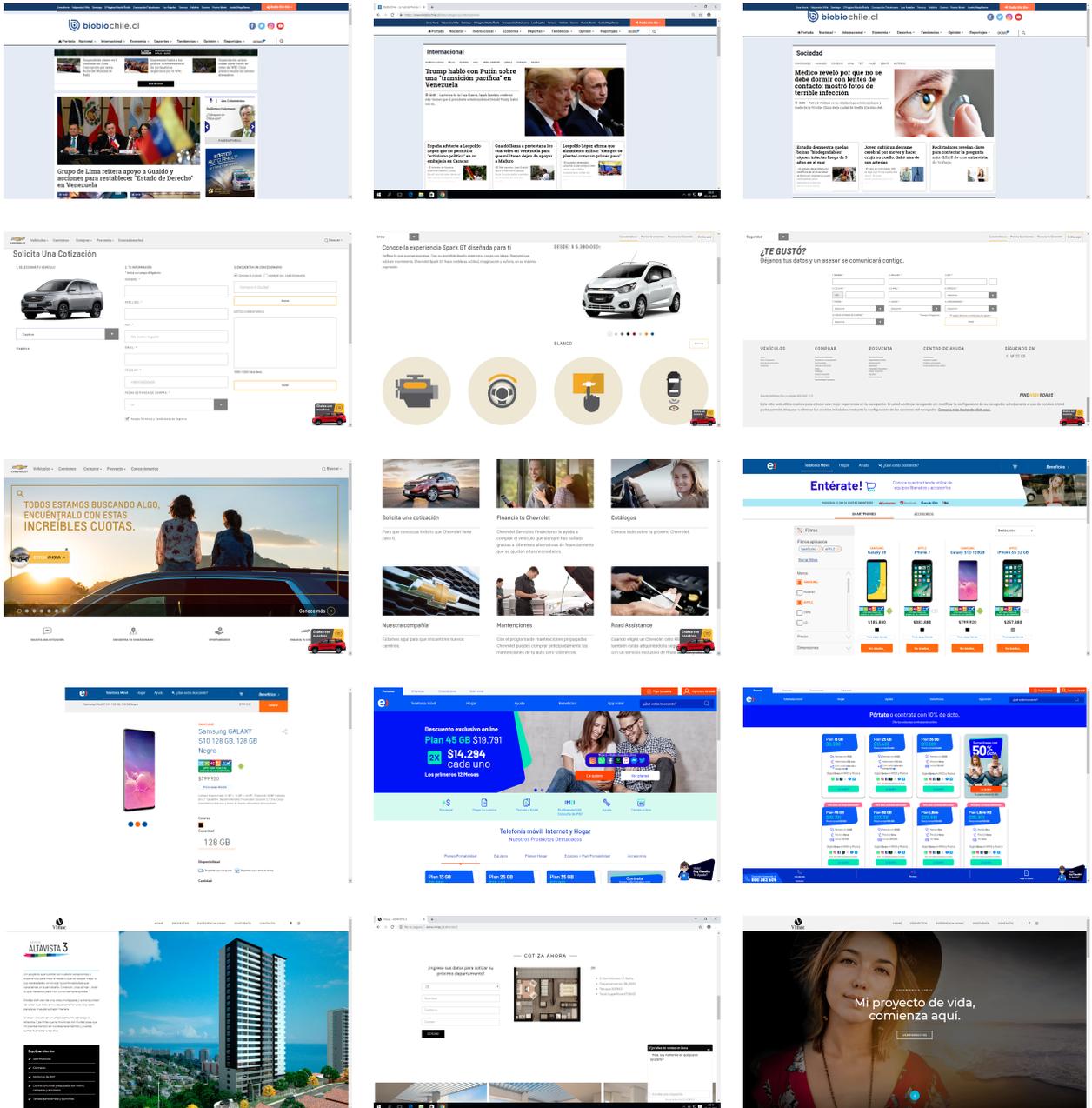


Figura 8.4: Conjunto de imágenes 2



Figura 8.5: Sala experimento