

5.5. Estudio del comportamiento atencional de usuarios en entornos virtuales

Ana Serrano¹
Belén Masiá²
Diego Gutiérrez³

1. Introducción

Los sistemas de realidad virtual nos proporcionan un nuevo medio que potencialmente tiene la capacidad de causar un gran impacto en nuestra sociedad. Las experiencias que nos ofrece esta nueva tecnología son inherentemente diferentes de la radio, la televisión, o el teatro, y eso nos ha permitido abrir nuevos horizontes y nuevas direcciones de investigación en áreas como la captura de contenido cinemático en realidad virtual, nuevas formas de interacción, o la generación y edición de contenido específico para esta plataforma. Sin embargo, la manera en la que este nuevo medio puede afectar al comportamiento atencional del usuario todavía es incierta, y, por tanto, no existen a día de hoy modelos estadísticos que consigan predecir con precisión este comportamiento, lo cual hace extremadamente complicado adecuar el diseño de nuevos contenidos a las expectativas de los usuarios.

Bajo este paradigma, emergen una serie de preguntas que debemos realizarnos a la hora de pensar en generar nuevo contenido para este entorno: ¿Cómo podemos diseñar escenarios con los que el usuario pueda interactuar de manera satisfactoria? ¿Podemos conducir la atención del usuario hacia los puntos de mayor relevancia de la escena? ¿Podemos predecir patrones de exploración comunes a los usuarios? Para bordar estas y otras cuestiones es crucial que entendamos cómo exploran los usuarios entornos virtuales. Un entendimiento detallado de la atención visual en realidad virtual no solo puede ayudar a responder las preguntas anteriores, sino que también puede ser extremadamente útil para futuras investigaciones en otros campos como por ejemplo interfaces de usuario, tecnología de seguimiento ocular, y otros aspectos clave en el diseño de sistemas de realidad virtual. Este documento recoge algunos de los aspectos más interesantes de nuestro trabajo, para información más detallada sobre los experimentos, y un análisis exhaustivo de nuestros datos, referimos al lector al trabajo completo desarrollado en colaboración con la universidad de Stanford (Sitzmann, 2018: 1633).

2. Captura del comportamiento de los usuarios

Para analizar el comportamiento de los usuarios en entornos inmersivos, se plantea una metodología escalonada, empezando por estímulos más sencillos, y creciendo en complejidad para construir sobre los conocimientos adquiridos. Para ello, en primer lugar, se quiere estudiar el comportamiento de los usuarios en escenas estáticas (en concreto, panoramas estereoscópicos omnidireccionales). Para ello, se ha realizado un extenso estudio donde se han capturado 1980 trayectorias del movimiento de la cabeza y la mirada de 169 personas diferentes explorando 22 entornos virtuales estáticos.

Para cada escena, se han explorado dos condiciones de visualización diferentes. Se plantea, por un lado, la visualización de las escenas en un dispositivo de realidad virtual, y, por otro lado, la visuali-

1] Universidad de Zaragoza, I3A. anase@unizar.es

2] Universidad de Zaragoza, I3A. bmasia@unizar.es

3] Universidad de Zaragoza, I3A. diegog@unizar.es

zación de las escenas en una pantalla tradicional. Las escenas en realidad virtual se muestran usando un dispositivo Oculus DK2, de manera que el usuario puede explorar el entorno de 360° con sus movimientos. El esquema de interacción en pantalla tradicional está diseñado de manera similar a reproductores de panoramas controlados por ratón en web, (por ejemplo, *YoutubeVR* o *Facebook360*), de manera que el usuario solo ve en la pantalla una porción de la escena, ya que el campo de visión limitado para imitar la reproducción en realidad virtual. Además, el usuario debe utilizar el ratón para explorar los 360 grados del panorama. En ambos casos, al usuario se le muestran una serie de escenas que debe explorar libremente, mientras se realiza un seguimiento ocular de los puntos a los que mira mientras realiza la exploración (fig. 1).



Fig. 1. Izquierda: Seguimiento ocular de los usuarios durante la exploración de una escena, donde cada uno de los puntos se corresponde con el punto al que un usuario estaba mirando en un determinado instante temporal, y las cajas asociadas se corresponden con la porción de la escena que el usuario estaba visualizando (campo de visión). Derecha: Mapa de calor (saliencia), resultante de agregar el comportamiento de todos los usuarios, que nos muestra las regiones de la escena que han resultado de mayor interés para todos ellos.

3. Análisis del comportamiento de los usuarios

Una vez capturados los datos de los usuarios, se procede a analizar su comportamiento durante la exploración de las escenas. Para ello, es importante identificar las áreas que le han resultado de interés (fijaciones), y las áreas que han sido descartadas (sacadas). Por lo general, los seres humanos no exploran una escena de manera estática, si no que los ojos se mueven buscando las partes más interesantes de dicha escena. Estos movimientos de búsqueda, se denominan sacadas (o movimientos sacádicos), y durante esta fase los ojos no extraen información visual de los estímulos que están observando. Tras las sacadas, se produce un momento de fijación ocular (o fijaciones), durante el que los ojos permanecen aproximadamente estáticos, y es en este momento cuando realmente se procesa la información de los estímulos visuales. Para averiguar los puntos de interés para el usuario, estamos interesados en las fijaciones, por lo que extraemos estos puntos de fijación de nuestros datos mediante un algoritmo basado en la dispersión y la duración de las fijaciones (Salvucci, 2000: 71). Una vez obtenidas las fijaciones para todos los usuarios, se pueden construir mapas de calor o saliencia para cada escena (fig. 2), que nos indican qué zonas del panorama han resultado de mayor interés a la mayoría de los usuarios (Le Meur, 2013: 251).

3.1 ¿Es posible extraer un patrón similar de exploración entre usuarios?

La primera cuestión que se aborda es intentar extraer un patrón similar de exploración de los usuarios para las diferentes escenas. Para ello, se analizan dos aspectos: Por un lado, el tiempo de exploración (fig. 3, izquierda), y, por otro lado, los mapas de calor medios resultantes de agregar todas las escenas presentadas (fig. 3, derecha).

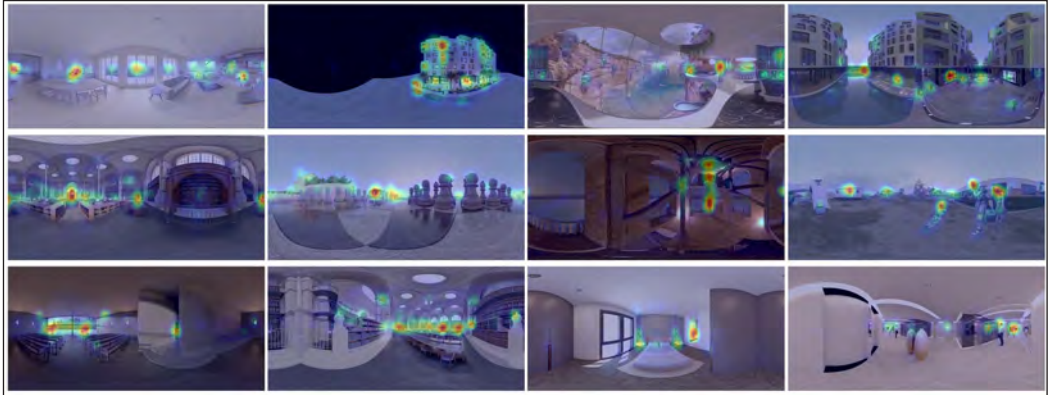


Fig. 2. Mapas de calor (saliencia) indicando, para un conjunto representativo de los panoramas, las regiones de las escenas que más atraen la atención de los usuarios. Cada mapa de calor se ha construido teniendo en cuenta el seguimiento ocular de todos los usuarios que han visualizado la escena. Adaptada de Sitzmann et al. (Sitzmann, 2018: 1633).

El tiempo de exploración, indica la cantidad de tiempo (en segundos) que le ha costado a un usuario recorrer toda la escena longitudinalmente, partiendo de un determinado punto. Podemos asegurar que los usuarios habrán recorrido la escena completa, en media, en unos 19 segundos. Además, esta exploración se realiza ligeramente más despacio cuando los usuarios visualizan las escenas en una pantalla tradicional (velocidad media de 23.63 grados/s frente a los 27.83 grados/s en realidad virtual). Esto podría deberse a que el mecanismo de interacción por ratón con la pantalla es menos natural y más laborioso que el mecanismo de interacción en realidad virtual, que consiste sencillamente en girar la cabeza.

Los mapas de calor medios se calculan como la media de todos los mapas de calor computados para cada una de las 22 escenas que se presentan a los usuarios. Nuestros datos indican que existe un *bias* entorno a la zona del ecuador, esto es, que los usuarios tienden a fijarse más en las regiones situadas a la altura de su línea de visión. Este fenómeno se produce tanto en la visualización en realidad virtual, como en la visualización en pantalla tradicional.

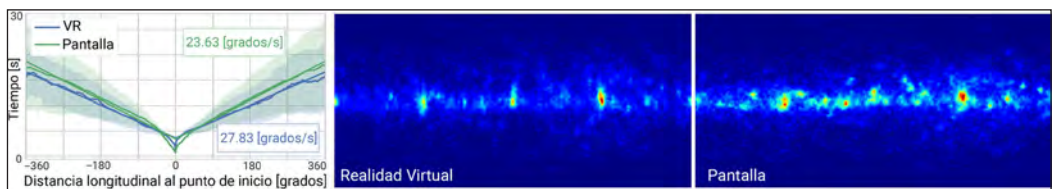


Fig. 3. Izquierda: Tiempo de exploración (segundos) de los usuarios en realidad virtual (VR), y en pantalla tradicional. Este tiempo se calcula como el tiempo medio hasta que los usuarios alcanzan a explorar cada uno de los offset longitudinales (grados) respecto al punto inicial donde se inició la exploración de la escena. Las líneas representan la velocidad de exploración media (en grados por segundo) para el total de los usuarios, mientras que el área de color contempla el abanico de diferentes velocidades de cada usuario. Derecha: Mapas de calor (saliencia) medios calculados con todas las escenas visualizadas en realidad virtual (izquierda) y en una pantalla tradicional (derecha). En estos mapas se puede apreciar un "bias en el ecuador", que indica que en ambas condiciones los usuarios prestan más atención a las regiones situadas a la altura de su línea de visión, la zona ecuatorial de la escena. Adaptada de Sitzmann et al. (Sitzmann, 2018: 1633).

3.3 ¿Es posible reconocer las zonas de interés sin necesidad de seguimiento ocular?

Conocer más acerca de las interrelaciones de los movimientos de la cabeza y los movimientos oculares es clave para el desarrollo de nuevas tecnologías en realidad virtual, por ejemplo, si somos capaces de

estimar las zonas de interés para el usuario a través de sus movimientos de cabeza, no necesitaremos costosos sistemas de seguimiento ocular.

Es bien conocido que la velocidad de la mirada es diferente cuando se producen, o no, fijaciones (Salvucci, 2000: 71), y, por tanto, podemos saber qué regiones despiertan interés al usuario, pero, ¿se produce el mismo efecto cuando nos fijamos en la velocidad de la cabeza? En la fig. 4 (izquierda) se puede ver la distribución de velocidades de la cabeza longitudinales para muestras correspondientes a fijaciones (morado), y a sacadas (verde). Nuestros datos muestran que los usuarios parecen comportarse de dos modos diferentes: atención y reorientación. Las fijaciones suceden en el modo de atención, cuando los usuarios están prestando atención a una zona en particular, mientras que movimientos rápidos a nuevas zonas de interés suceden en el modo de reorientación. Los usuarios mueven su cabeza significativamente más rápido cuando no están interesados en la zona que están mirando (50.13 grados/s, frente a los 19.59 grados/s cuando se producen fijaciones). Ser capaces de detectar estos dos modos en tiempo real puede ser muy útil para el desarrollo de nuevas técnicas de interacción en tiempo real.

Además, nuestros datos parecen indicar que también puede distinguirse una diferenciación entre estos dos modos de exploración en el offset o desviación de los ojos con respecto a su posición central (fig. 4, derecha). Los usuarios mantienen sus ojos en una posición más natural (más cercanos a la posición central) cuando están en el modo de atención, mientras que, por otro lado, se producen excentricidades más acusadas hacia los lados cuando los usuarios buscan nuevas zonas que explorar (modo de reorientación).

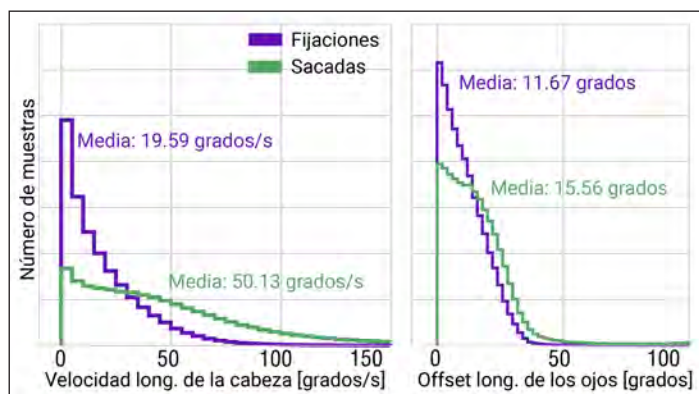


Fig. 4: Distribuciones de la velocidad longitudinal de la cabeza (izquierda), y el offset de los ojos con respecto a su orientación central (derecha) según si el usuario está realizando sacadas, en verde, o fijaciones, en morado. Adaptada de Sitzmann et al. (Sitzmann, 2018: 1633).

4. Conclusiones y trabajo futuro

En este documento hemos recogido algunos de los aspectos más interesantes nuestro trabajo en colaboración con la universidad de Stanford (para un análisis detallado, referimos al lector al trabajo completo (Sitzmann, 2018: 1633)). En concreto, hemos tratado de dar respuesta a algunas preguntas claves sobre el comportamiento de los usuarios a través del análisis de 1980 trayectorias del movimiento de la cabeza y la mirada de 169 personas diferentes explorando 22 entornos virtuales estáticos. Conocer más acerca del comportamiento de los usuarios en entornos inmersivos es crucial para desarrollar nuevos mecanismos de interacción y sistemas de visualización (Serrano, 2019), así como diseñar nuevos contenidos, incluyendo, por ejemplo, contenidos cinematográficos para realidad virtual (Serrano, 2017: 47).

Existen muchas vías potenciales de trabajo futuro. Por ejemplo, se podría extender el análisis del comportamiento de los usuarios a escenarios interactivos o dinámicos (en lugar de solo estáticos), a estímulos visuales de mayor complejidad, o a entornos con sonidos direccional. Otra posible vía de investigación futura radicaría en diseñar modelos predictivos que fuesen capaces de adaptarse a los

diferentes patrones de visionado de diferentes usuarios, y pudiesen predecir los recorridos visuales particulares de cada usuario dada una escena determinada. En esta línea, nuestros datos capturados podrían servir de punto de partida para entrenar dichos modelos.

Bibliografía

- LE MEUR, O., y BACCINO, T. (2013). "Methods for comparing scanpaths and saliency maps: strengths and weaknesses". *Behavior research methods*, Berlín: Springer, vol. 45(1), pp. 251-266.
- SALVUCCI, D. D., y GOLDBERG, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols". En *Proceedings of the 2000 symposium on Eye tracking research & applications*, New York: ACM, pp. 71-78.
- SERRANO, A., SITZMANN, V., RUIZ-BORAU, J., WETZSTEIN, G., GUTIERREZ, D., Y MASIA, B. (2017). "Movie editing and cognitive event segmentation in virtual reality video". *Transactions on Graphics (TOG)*, Anaheim: ACM, vol. 36(4), pp. 47-59.
- SERRANO, A., KIM, I., CHEN, Z., DIVERDI, S., GUTIERREZ, D., HERTZMANN, A., Y MASIA, B. (2019). "Motion parallax for 360 RGBD video". *IEEE transactions on visualization and computer graphics*, Osaka: IEEE, vol. 25(4).
- SITZMANN, V., SERRANO, A., PAVEL, A., AGRAWALA, M., GUTIERREZ, D., MASIA, B., Y WETZSTEIN, G. (2018). "Saliency in VR: How do people explore virtual environments?". *IEEE transactions on visualization and computer graphics*, Reutlingen: IEEE, vol. 24(4), pp. 1633-1642.