

Diseño de audiojuegos: estado del arte y desarrollo de un prototipo

Audio game design: state of the art and prototype development

Presentación: 4 y 5 de Octubre de 2022

Doctorando:

Guillermo Gilberto

Centro de Investigación y Transferencia en Acústica (CINTRA), Universidad Tecnológica Nacional (UTN) y Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET) - Argentina
lgilberto@frc.utn.edu.ar

Director:

Cristian García Bauza

Codirectores:

Fernando Bermejo - Fabián Tommasini

Resumen

La industria de los videojuegos ha tenido un crecimiento importante y sostenido durante las últimas décadas, sin embargo sus condiciones de accesibilidad son un aspecto pendiente. A pesar de los beneficios demostrados en múltiples dimensiones del ser humano (social, cognitivo, psicomotriz) existen actualmente muy pocos desarrollos para personas con discapacidad. Por su parte, los audiojuegos son un tipo de juegos computacionales que no disponen de claves visuales, por lo tanto accesibles a personas con discapacidad visual. Este proyecto doctoral busca contribuir al avance del conocimiento y desarrollo de este tipo de tecnologías. Se espera que favorezca la generación de aplicaciones para el entretenimiento, entrenamiento de audición espacial e inclusión de personas con diferentes capacidades visuales. Se presentan en este trabajo antecedentes de este tipo de tecnología y el diseño de un audiojuego de realidad virtual para el entrenamiento de audición espacial con resultados preliminares de estudios experimentales realizados con él.

Palabras clave: audiojuegos, interfaces humano-máquina, audición espacial, discapacidad visual.

Abstract

Videogame industry has undergone a very important and sustained growth during the last decades, however, its accessibility conditions are a pending issue. Despite the evidence of the benefits of these games in multiple dimensions of the human being (social, cognitive, psychomotor) there are currently very few developments for people with disabilities. In this context, audio games are a type of computer games that do not have visual cues, therefore accessible to visually impaired people. This doctoral project seeks to contribute to the advancement of knowledge and development of this type of technology. It is expected that this project will favor the generation of applications for entertainment, spatial hearing training and inclusion of people with different visual abilities. This paper presents background information on this type of technology and the design of a virtual reality audiogame for spatial hearing training with preliminary results of experimental studies carried out with it.

Keywords: audiogames, human-machine interfaces, spatial audition, visual impairment.

Introducción

Los videojuegos se han convertido en las últimas décadas en una parte muy importante de la cultura de la sociedad, siendo hoy en día una de las principales tecnologías de entretenimiento preferidas por millones de personas. Si bien su mayor utilización se da en contextos de ocio y recreación, diversos estudios señalan importantes beneficios cuando se los usa como recurso educativo, de entrenamiento y rehabilitador (Fuchs et al., 2014), entre otras. Una de sus mayores fortalezas está en su poder motivacional; el desafío continuo y el rol activo que toma el jugador son muy favorables para el aprendizaje y entrenamiento de variadas habilidades cognitivas, sociales y psicomotrices (Graells, 2005: 80-89). Los avances tecnológicos de los videojuegos son notables y fácilmente observables. Desde el punto de vista de hardware existen mejoras continuas en el poder de cómputo de placas de video e interfaces de juego cada vez más complejas con múltiples sensores. Cascos de realidad virtual, aumentada y mixta, acelerómetros, cámaras, conexiones inalámbricas, estimuladores hápticos, etc. son algunos de estos elementos que han estado modificando el panorama de las interfaces año tras año. Por su parte, la irrupción en la vida cotidiana de teléfonos celulares, tablets, *smarts* TV y diferentes dispositivos capaces de ejecutar aplicaciones multiplican su popularidad. La conectividad entre jugadores también muestra el progreso de estas tecnologías con los denominados *Massive Multiplayer Online Games* (MMO) y una creciente comunidad de jugadores que ya en 2017 se estimaba en 2210 mil millones (un tercio de la población mundial) (Nesteriuk, 2018: 338-352).

A pesar de esta evolución y éxito creciente, la accesibilidad en videojuegos es todavía una asignatura pendiente (Pérez-Castilla Alvarez, 2012). Considerando exclusivamente la población con discapacidad visual, el análisis se centra sobre los denominados audiojuegos. Estos juegos de computadora prescinden de interfaces visuales e implementan un tipo de interacción basada principalmente en claves auditivas, posibilitando el acceso a personas ciegas. Tanto la mecánica de juego como el ambiente virtual y cualquier información (puntajes, instrucciones, tiempo, etc.) son presentadas al jugador mediante sonidos. En algunos casos también incorporan claves hápticas, aunque su utilización suele ser complementaria a la estimulación sonora.

Más allá de su aporte al ocio y entretenimiento, los audiojuegos tienen potencial para el entrenamiento de la audición espacial (habilidad para localizar sonidos en el espacio). Esto cobra especial interés ya que esto puede contribuir y mediar sobre una de las limitaciones más severas que impone la ceguera, como lo es la restricción de la movilidad independiente. El desarrollo de habilidades de Orientación y Movilidad en personas con discapacidad visual requiere de un cuidadoso y extensivo aprendizaje. El entrenamiento sistemático de habilidades auditivas espacial es uno de los pilares de tales programas de aprendizaje. La audición entrenada permite que el individuo logre un estado de alerta altamente sensible para saber qué ocurre en el entorno y le posibilita desarrollar habilidades de orientación espacial para percibir y sortear de mejor manera obstáculos presentes en su camino. Los escasos estudios existentes en la temática indican que las personas ciegas logran adquirir adecuados perceptos espaciales y pueden transferir el aprendizaje de las habilidades que logran en la plataformas virtuales, tales como los audiojuegos, a situaciones de la vida real (Merabet et al., 2012; Allain et al., 2015; Balan et al., 2015).

Complementariamente, este tipo de juegos ofrece a personas con visión una opción alternativa a los videojuegos. Un conveniente diseño de audiojuegos tiene un valor agregado para estos usuarios ya que implica el desafío novedoso de jugar sin visión, sentido dominante en su desempeño cotidiano. En este trabajo presentamos una breve revisión del estado del arte de los audiojuegos y el diseño de un audiojuego de realidad virtual para el entrenamiento de audición espacial. Además se muestra la implementación de una prueba de localización sonora desarrollada *ad hoc* para estudiar su funcionamiento y los resultados experimentales preliminares con personas con visión.

Desarrollo

Un repaso por el Estado del Arte de los Audiojuegos

En un estudio realizado con 47 audiojuegos en el marco de este proyecto doctoral (Gilberto et al., 2021), se encontró que el desarrollo de audiojuegos en la actualidad no cuenta con grandes aportes de las empresas líderes en la industria del entretenimiento digital. Por el contrario, se funda con contribuciones que vienen desde pequeños estudios de desarrollo independientes, proyectos académicos en la temática y/o a partir de iniciativas individuales

de programadores entusiastas. Los proyectos provienen mayoritariamente de Estados Unidos, seguido por algunos países europeos. En Latinoamérica se encuentra la mayoría de desarrollos en Brasil y Chile. A diferencia de lo que ocurre con los videojuegos, el origen de los audiojuegos es clave ya que esto determina el idioma principal en el que se propone la interacción. Dado que gran caudal de la información se conduce a través del lenguaje hablado, resulta en algunos casos inhibitoria la jugabilidad si no se maneja el idioma propuesto. En este sentido, se encontró que sólo el 30% contaba con la opción de ejecutarse en idioma español.

En este mismo estudio se observó un crecimiento en la cantidad de títulos lanzados que coincide con los años de explosión de teléfonos inteligentes, esto es 2012-2014. Los tipos de juegos encontrados cubren un amplio espectro: acción, aventura, agilidad mental, deportivos, musicales, etc. Es destacable que en este estudio no se encontraron audiojuegos para consolas, aunque sí para computadoras personales, tablets y teléfonos inteligentes. Se encontró en este estudio también que el 85% de los juegos tenían como principal finalidad el entretenimiento, y sólo un 15% introducía además intenciones de rehabilitación, entrenamiento, ejercicio físico o alguna finalidad educativa. Un análisis detallado sobre las interfaces de sonido arrojó que la espacialización sonora era básica o nula en la mayoría de los casos. Esto supone un detrimento en el poder de interacción, un desperdicio de las posibilidades que ofrece este principal canal de comunicación con el jugador. Por su parte, los medios de interacción de entrada de los audiojuegos son mayoritariamente teclado y mouse (55%), seguidos de interfaces táctiles (25%). Finalmente se observó en esta muestra que son muy escasos los juegos que permiten la interacción con otros jugadores.

Audiojuego Sombras y Sonidos

Consideraciones generales: A partir del estudio de las características principales y particularidades del campo de desarrollo dentro de los audiojuegos se establecieron algunos lineamientos de diseño exploratorios. Estas premisas iniciales tomaron en cuenta consideraciones en torno a interfaces hombre máquina y postulados de Cs. Cognitivas referidas a cognición corporizada para dar forma a un modelo de interactividad directa, natural e intuitiva. Si bien se realizaron avances en varios diseños, este documento focaliza sobre uno de los desarrollos realizado para teléfonos móviles y con la utilización de auriculares.

Objetivo: crear un audiojuego que sea al mismo tiempo desafiante, motivante para jugarlo y que tenga potencialidades para el entrenamiento de audición espacial en el plano horizontal.

Herramientas de diseño: se utilizó el motor de desarrollo de contenido en tiempo real *Unity*. Esta plataforma es ampliamente reconocida y se usa principalmente para el desarrollo de videojuegos, pero también en arquitectura, cine, en la industria automotriz, etc. Cuenta con la posibilidad de adquirir una licencia gratuita para ingresos o financiamientos menores a los 100 mil dólares anuales. El poder interactivo de este motor es muy interesante aunque sus características para manipular audio son limitadas para construir un audiojuego. Por esta razón se evaluaron de diferentes plugins y middlewares capaces de acoplarse a este motor y dar un mejor soporte de audio interactivo. Se definió utilizar el plugin de la empresa *Valve Software* denominado *Steam Audio*. Éste ofrece una solución de audio completa que integra simulación de entornos y utiliza modelos de función de transferencia de cabeza y torso genéricas (HRTF) que mejoran significativamente la inmersión.

Implementación: el diseño se realizó pensando en teléfonos celulares de gama media que cuenten con el sensor giroscópico (teléfono de referencia: *Samsung J3+*). Este sensor sumado al visor sujetador tipo *VR Case* (Figura 1-a) que mantiene solidariamente unido el teléfono a la cabeza del jugador permite introducir la perspectiva de escucha de éste dentro de un entorno virtual y realimentar el sistema constantemente. Completan el esquema de hardware necesario auriculares y un joystick inalámbrico *Bluetooth*. El desarrollo tanto de la idea como la mecánica general de juego se realizó dentro del equipo de investigación conformado por diferentes profesionales del área de Ingeniería y Psicología del CINTRA y del Media.Lab. Se buscó una interactividad intuitiva para el jugador, con libertad para girar la cabeza y/o el cuerpo, intentando de esta manera reproducir los lazos sensoriomotores implicados en la escucha cotidiana (particularmente los movimientos necesarios para localizar sonidos). También la propuesta debía invitar a una atención auditiva activa y constante, requiriéndole al jugador recurrentemente la utilización de mecanismos de identificación y localización sonora. El ambiente virtual se pactó de complejidad media en cuanto a los enmascaramientos sonoros producidos. Esto último refiere a qué tan congestionado de otros sonidos se encuentra el ambiente: demasiados estímulos pueden generar desagrado, fatiga y ansiedad, mientras que un

entorno muy simple reduce el desafío y la motivación. La ejecución técnica estuvo a cargo del doctorando y participaron además 2 becarios alumnos de investigación mediante trabajos remotos (debido a las restricciones de la pandemia de COVID-19) y sincronizaciones mediante repositorios tipo *git*. La codificación de scripts dentro de *Unity* se realizó en lenguaje C#, que es el más utilizado en la plataforma y con más referencias externas.

Descripción del audiojuego: el audiojuego se denominó Sombras y Sonidos (SyS) y toma lugar en un terreno virtual amplio, nocturno y al aire libre. La tarea del jugador es identificar, localizar y recolectar en un orden determinado diez animales distintos que se encuentran distribuidos en el terreno y sonando continuamente. SyS no cuenta con ningún tipo de clave visual, ya que la pantalla del teléfono permanece apagada durante toda la partida. Una vez iniciado el juego, una voz guía indica cuál es el animal que debe ser hallado, acompañado de una breve reproducción del sonido característico de dicho animal. El jugador debe lograr acercarse hasta él realizando movimientos rotacionales con su cabeza y/o cuerpo para panear la escena y moviendo una palanca analógica del joystick para trasladarse. Cuando se encuentra lo suficientemente cerca del animal de turno, debe presionar un botón del joystick para recolectarlo. En ese momento el animal deja de sonar y se escucha un sonido que confirma la recolección; automáticamente después la voz guía indica cual es el próximo animal a encontrar.

El juego finaliza cuando se encuentran 10 animales (caballo, gallinas, cabras, vaca, ovejas, gallo, burro, pavo, cerdo y patos); el orden de búsqueda de los animales y su posición en el espacio virtual cambia en cada juego. El terreno virtual es de forma cuadrada y el traslado del personaje en él conlleva un sonido de pasos pisando gramilla. Si se transita cerca de los límites del terreno se escucha que los pasos del jugador pisan agua, y si se llega efectivamente al límite del terreno se escucha un sonido parecido al rebote de una pelota indicando que no se puede ir más allá. Finalmente, en caso de olvido, se cuenta con un botón que hace que la voz guía le recuerde cuál es el animal que tiene que encontrar y su sonido característico.

Pruebas subjetivas

Se realizó un estudio prueba para evaluar las potencialidades de SyS para el entrenamiento de audición espacial. En esta dirección, primeramente se realizó una revisión de 22 artículos científicos que ensayaban sobre diferentes aspectos de este tipo de entrenamiento. Allí se detectaron puntos relevantes a tener en cuenta como métodos de entrenamiento y de virtualización sonora, utilización de HRTFs, muestra de participantes, estímulos sonoros utilizados, etc. y se elaboraron las pautas de los ensayos con el audiojuego. Se decidió en esta primera etapa del doctorado trabajar con personas con visión, antes de probar los desarrollos con personas con discapacidad visual.

Se diseñó una prueba de localización sonora para ejecutarse con dos grupos de personas: un grupo con entrenamiento (GE) y grupo control (GC). GE realizaba las pruebas previo y posterior al entrenamiento con el audiojuego y GC realizaba idénticas pruebas pero sin realizar el entrenamiento. Las pruebas consistían en localizar una serie de estímulos sonoros presentados alrededor del plano horizontal de los participantes, utilizando idénticas interfaces que en el audiojuego (Figura 1-a). Considerando algunas diferencias encontradas en la literatura se eligió testear la localización de sonidos con tres estímulos diferentes: 1- ruido rosa de banda ancha (20 Hz a 20 kHz) en pulso único con una duración de 500 ms (RRU), 2- ruido rosa de banda ancha en tren de pulsos con igual duración del pulso y de silencio: 500 ms (RRL), y 3- sonido del cacareo de una gallina (GAL). Los dos primeros son ruidos de laboratorio muy utilizados en este tipo de trabajo; tienen la ventaja de estimular con igual cantidad de energía todas las bandas de octava percibibles por un ser humano con audición normal. El último es un ruido que consideramos más ecológico y es uno de los animales que aparece en el juego.

Un total de 18 personas participaron de los ensayos. Se asignaron aleatoriamente 9 personas al grupo experimental (GE) y 9 al grupo control (GC). Todos los participantes, luego de recibir la consigna y una breve familiarización con los materiales y la prueba, debían resolver la prueba en tres etapas que duraban aproximadamente 15 minutos cada una y con un descanso de 5 minutos entre ellas. Primeramente lo hacían con el estímulo RRU, luego con el RRL y finalmente con GAL. La tarea del participante consistía en rotar la cabeza y/o el torso cada vez que escuchaba un estímulo hasta enfrentarlo con su rostro y luego presionar un botón del joystick para indicar su respuesta. Cada vez que respondía debía volver a su posición inicial ayudado por un sistema de sonidos implementado para tal fin. Los sonidos presentados se ubicaban a la altura del rostro y se distribuían aleatoriamente en 12 posiciones: 0° (en frente), ±30°, ±60°, ±90°, ±120°, ±150° y 180° (atrás). En cada una de estas posiciones el estímulo aparecía 4 veces, totalizando

144 ensayos para una sesión de prueba completa. Las pruebas se tomaron en recintos con bajo ruido de fondo (cámara acústica del CINTRA o en sala de grabación con tratamiento acústico).

Luego del pretest, sólo los participantes del GE pasaban a la etapa de entrenamiento. El entrenamiento se realizaba en 7 encuentros en diferentes días, en un lapso no mayor a 15 días luego de ejecutado el pretest. En cada una de estas sesiones el participante debía jugar SyS durante 30 minutos o completar 2 juegos separados de un breve descanso. El primer día de entrenamiento, con el objetivo de familiarizarse con la tarea, el participante debía completar también una escena de familiarización que consistía en la búsqueda de ocho balizas (sonido de agudo de *beep*) en un terreno similar al de SyS, pero sin ruido ambiente. Esta escena tenía la intención de presentar al jugador todos los elementos del audiojuego de manera simplificada, ya que en las pruebas pilotos se detectó que el escenario del juego y el manejo de las interfaces podría ser algo complejo la primera vez que se las utilizaba. Los entrenamientos se realizaban en los domicilios de los participantes o en las mismas salas de las pruebas. Finalmente, luego de un lapso de entre 10 y 15 días del pretest, ambos grupos realizaban el postest en idénticas condiciones. Una entrevista con aspectos relevantes a la experiencia de juego fue realizada con el GE una vez terminado el esquema de pruebas.

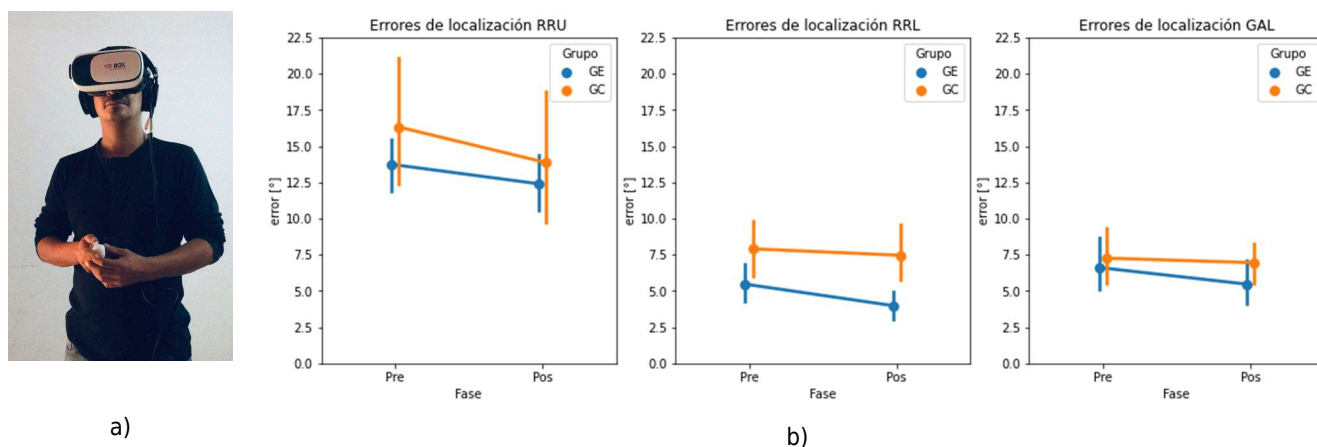


Figura 1: a) participante utilizando las interfaces de las pruebas y audiojuego: auriculares, visor portador de teléfono celular y joystick de conexión inalámbrica; b) errores medios angulares en grados sexagesimales e índices de confianza del 95% para la

Resultados

Para evaluar los resultados de las pruebas experimentales se realizó en primera instancia un preprocesamiento de los datos relevados. Esto se realizó debido a que se observó una gran variabilidad en las respuestas de los participantes durante las pruebas. Luego de este filtrado se procesaron los datos mediante el software *JupyterLab* del *Anaconda* para ciencia de datos, y se efectuó un análisis descriptivo y gráfico que se muestran a continuación.

El preprocesamiento constó de dos fases. En primer lugar se detectaron errores de tipo *reversal*, que son equivocaciones propias de entornos auditivos virtuales cuando se utilizan HRTF genéricas. Un *reversal* se da cuando un participante percibe una fuente sonora de manera "espejada" en el hemisferio contrario al que en realidad se presenta, siendo el ejemplo más común percibir una fuente que está a 0° (en frente) a 180° (atrás) y viceversa. En segundo lugar se realizó un filtrado extra quitando valores atípicos mediante el método de distancia intercuartil de las distribuciones. Estos últimos errores se explican a partir de equivocaciones en la pulsación de los botones o desconcentraciones de los participantes durante la toma de pruebas.

A partir de allí se calcularon las medias de los errores de localización y se obtuvieron las gráficas de la Figura 1-b. Si bien previo al filtrado de datos se observó una tendencia a una reducción mayor del error en el grupo entrenado para todos los estímulos, luego del preprocesamiento los indicadores se modificaron. La tendencia se mantuvo para los estímulos prolongados (RRL y GAL) pero se perdió para el estímulo breve (RRU), observándose una mejoría superior en GC vs. GE. Se observa además que en general para este estímulo los errores son mayores, con medias cercanas a los 15°. Esto tiene que ver con que la información que percibe el sujeto en este estímulo es mucho menor y además se encuentra con menos posibilidades de interacción dinámica. Por otra parte, si bien SyS cuenta con un

entorno sonoro muy rico, los objetivos de juego se pautan en torno a estímulos de características prolongadas (voces de animales). Posiblemente el entrenamiento para la localización de sonidos breves necesite tratamiento especial dentro del juego. Finalmente, en el proceso de filtrado comentado anteriormente, encontramos que para RRU la cantidad de errores por *reversals* desciende un 24,7% en el postest vs pretest para GE, y por el contrario, en GC se produce un aumento del 3%. Esto podría demostrar cierto aprendizaje en la interacción con el entorno virtual exclusivamente, indicando que los participantes logran adaptarse en alguna medida a la HRTF genérica utilizada.

En referencia a los otros dos estímulos, se observa una mayor tendencia a la mejoría para el estímulo RRL. Este resultado es interesante y requiere un análisis más profundo, sobre todo por la gran dispersión que se observa. Algunas conjeturas preliminares sobre estos resultados recaen en propiedades acústicas del estímulo de GAL: un espectro de frecuencia acotado y un nivel sonoro oscilante, propio del cacareo de una gallina. El análisis se vuelve aún más complejo si se toma en cuenta cómo varios participantes mencionaron explícitamente que localizar la gallina era más "fácil". Todos estos resultados están siendo estudiados en profundidad actualmente por el doctorando y el equipo de investigación mediante análisis estadísticos inferenciales.

Conclusiones

En este documento se mostraron algunos de los avances de los últimos 2 años dentro del proyecto doctoral. Aunque gran parte de este tiempo sucedió en aislamiento por la pandemia de COVID-19 se pudieron concretar progresos sustanciales como la revisión del estado del arte de los audiojuegos, el diseño e implementación de un audiojuego para el entrenamiento de audición espacial, una revisión sobre metodologías utilizadas en investigaciones de localización sonora y el diseño de las pruebas experimentales para testear el funcionamiento de la audiojuego diseñado.

Una vez levantadas las restricciones de aislamiento se pudieron tomar las pruebas experimentales con 18 personas con visión. Los primeros resultados son promisorios, indicando un conveniente desempeño de la herramienta diseñada. Los sujetos entrenados con el audiojuego evidencian una reducción en los errores de localización sonora superior a la del grupo que no entrenó con el audiojuego. En esta porción del doctorado queda por delante un análisis detallado de los datos obtenidos en las pruebas y en las entrevistas realizadas para completar las conclusiones de esta etapa y realizar las pruebas experimentales con personas con discapacidad visual.

Referencias

- Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R. & Kybartas, B. (2015, March). An audio game for training navigation skills of blind children. *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, 1-4. IEEE.
- Balan, O., Moldoveanu, A., & Moldoveanu, F. (2015). Navigational audio games: an effective approach toward improving spatial contextual learning for blind people. *International Journal on Disability and Human Development*, 14(2), 109-118.
- Fuchs, M., Fizek, S., Ruffino, P., & Schrape, N. (2014). Rethinking gamification. Meson press.
- Gilberto, L.G., Arellano, P., Miloro, E., Bermejo, F. & Hüg, M. (2021). Disponibilidad y características de juegos electrónicos accesibles a personas ciegas. *XVIII Reunión Nacional VII Reunión Internacional de la de la Asociación Argentina de Ciencias del Comportamiento*, Mar del Plata, Argentina, 22 - 24 de septiembre.
- Graells, P. M. (2005). "Nuevos entornos, nuevos modelos didácticos", *Cuadernos de pedagogía*, 363, 80-89.
- Merabet, L. B., Connors, E. C., Halko, M. A., & Sánchez, J. (2012). Teaching the blind to find their way by playing video games.
- Nesteriuk, S. (2018). Audiogames: Accessibility and inclusion in digital entertainment, *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, 338-352. Springer, Cham.
- Pérez-Castilla Alvarez, L. (2012). Buenas prácticas de accesibilidad en videojuegos. Madrid, IMSERSO.