

# Localización de sonidos en entorno abierto por personas ciegas



Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Fernando Brusola

Universitat Politècnica de València (Spain)  
islenlen@upv.es

PACS: 43.66.Qp, 43.66.Yw

## Resumen

El artículo describe los resultados obtenidos en la localización de sonidos virtuales mediante un dispositivo detector de objetos en entorno de laboratorio y en entorno abierto con personas invidentes.

En las pruebas han participado dos grupos de diez personas invidentes. Los participantes tenían edades comprendidas entre 25 y 60 años. El dispositivo detector de obstáculos reproducía 64 sonidos en un azimut de 64° para una distancia de entre 0,5m hasta 5m en un plano horizontal al nivel de los ojos. Los sonidos virtuales se reproducían mediante auriculares.

Para el experimento se ha utilizado un sonido grabado que varía entre -110dB y -23dB con frecuencia entre 86Hz hasta 18.604Hz. El sonido se reprodujo a una frecuencia de 44.1kHz y 32 bits, y con una duración de 0.04641s.

Los resultados nos muestran que las personas invidentes son capaces de localizar objetos mediante sonidos simulados virtualmente y detectarlos con gran precisión.

## Abstract

The paper describes the results of virtual sound localization via an object detector in laboratory setting and open environment with blind people.

In the tests have been involved two groups of ten blind people. Participants were aged between 25 and 60. The obstacle detection device reproduced 64 sounds at an azimuth of 64° for a distance of between 0.5 m to 5m in a horizontal plane to the level of the eyes. Virtual sounds were reproduced through headphones.

For the experiment we used a recorded sound which varies between -110dB and -23dB with frequency between 86Hz to 18.604Hz. The sound was reproduced 44.1kHz frequency and 32 bits, and a duration of 0.04641s.

The results show that blind people are able to locate objects by virtually simulated sounds and detect them with high accuracy.

## 1 Introducción

Los seres humanos tenemos una gran capacidad de localización de una la fuente de sonido, de hecho es uno de los factores más importantes para la supervivencia humana. Esta capacidad nos ayuda a percibir, detectar peligros y diferenciar entre los cambios de las condiciones climáticas, ambientales, etc. La localización de la fuente de sonido es la «ley o regla por la cual la ubicación de un evento auditivo se relaciona con un determinado atributo o atributos de un evento de sonido» [1]. La localización de una fuente de sonido viene caracterizada por los siguientes parámetros: la diferencia de Tiempo Interaural (ITD), la Diferencia de nivel de Interaural (ILD) así

como la geometría del oreja [2]. El uso de las funciones Head Related Transfer Functions (HRTF) para la localización de fuentes de Sonido Virtuales es imprescindible ya que representan las características de transferencia de la fuente de sonido en un campo libre para el oído externo del oyente.

La localización de sonidos ha sido siempre un tema de interés en acústica. Hasta el presente, se ha estudiado la localización de sonidos en dirección, llamado lateralización, para distancias cortas de un metro [1]. La mayoría de los experimentos se han producido dentro de las cámaras anecoicas o salas con reducción de ruido [3], [4]. También, el objetivo de los experimentos en localización

de sonidos es la de analizar y validar la influencia de los parámetros acústicos, como la Diferencia Interaural de Tiempo [5], la Diferencia Interaural de Intensidad para las altas Frecuencias, el cono de confusión, el efecto de reverberación, etc.

En la actualidad existe poca investigación con respecto a la localización de sonidos para personas con discapacidad visual.

El presente trabajo describe los métodos y resultados de la localización de las fuentes de sonido virtual por personas ciegas mediante el uso de las funciones (HRTF).

## 2 Descripción de los equipos

El principal objetivo del proyecto es llegar a transmitir lo que un ser humano detecta cuando está caminando de forma consciente e inconsciente y que le permite poder deambular sin interceptar los objetos que va encontrando por su camino. La información que percibe el ser humano es muy amplia como puede ser la profundidad de campo, acimut, altura, etc... Toda la información hay que clasificarla previamente, no sin dificultad, para decidir cuál es la información más importante que hay transmitir para poder definir un espacio de forma unívoca mediante la representación de una plano sonoro de una escena que queremos representar. La representación del espacio mediante sonidos no es trivial y hay que definir claramente como representar los objetos de una escena. Una de las características que se aprovechan en este proyecto es la capacidad que tiene en sistema auditivo humano de localizar varias fuentes de sonido al mismo tiempo [6]. También cabe destacar la capacidad de discriminación y combinación del oído humano y su capacidad de combinarlas por clases de señales y frecuencias.

En la actualidad los sistemas de navegación existentes se basan en diferentes tecnologías de captación como son el láser, sistemas basados en estereovisión, dispositivos basados en geolocalización por satélite GPS, incluso sistemas basados en ultrasonidos. Todos



Figura 2. Imagen real con los elementos del sistema: gafas (elemento sensor), auriculares y FPGA.

estos sistemas requieren de una gran complejidad computacional del algoritmo de interpretación, lo que conlleva a requerir sistemas de procesamiento de muy alto coste. Algunos de estos sistemas transmiten la información precisa vía voz, sin llegar a ser en tiempo real con el inconveniente de confundirse con la voz humana.

La solución adoptada ha sido desarrollar un sistema cognitivo, que resuelve los principales problemas, el objetivo es transmitir al usuario la información del espacio que le rodea mediante un sistema de audio [7]. En la figura 1 se muestra de forma esquemática el sistema sensorial auditivo desarrollado.

El dispositivo sensor está basado un CMOS láser que emite una luz láser con una definición de 64 puntos con una diferencia de un grado de separación ente cada punto, abarcando una angulación total de 64° la distancia máxima a alcanzar son 5 metros. Este dispositivo nos permite obtener la distancia y azimut exacto donde está ubicado un objeto. Una vez conocida la distancia y ángulo se procesa mediante un algoritmito que atribuye a esa posición una secuencia de sonidos virtuales que permiten el posicionamiento univoco del objeto.

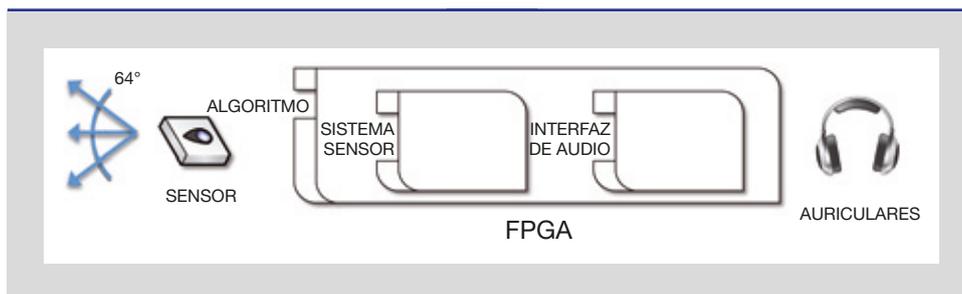


Figura 1. Representación esquemática del Sistema.

## 2.1 Detección de distancias

Una vez definido todo el proceso y generados los algoritmos necesarios para el funcionamiento óptimo del sistema, el principal objetivo era integrar el dispositivo sensor de entrada en un dispositivo que fuera lo más cómodo posible y que el usuario en este caso invidente lo aceptara como un elemento más en su vida cotidiana. La solución adoptada fue integrar el dispositivo sensor compuesto por componentes ópticos, componentes electrónicos y un CMOS láser en unas gafas con apariencia similar a unas gafas de sol que la mayoría de personas invidentes suelen utilizar. El componente CMOS láser está situado justo en el centro de las gafas en la posición que correspondería a la montura justo encima de la nariz lo que proporciona una posición totalmente centrada con respecto a la lateralidad del usuario, se utilizó un polímero para simular el cristal de las gafas que no interfiere con la luz láser y tanto las pruebas realizadas para la medición de distancias con el polímero como sin polímero dieron los mismo resultados.

El sensor permite obtener las medidas de los objetos que tiene alrededor mediante los 64 pixeles que tiene el dispositivo láser que permite calcular con muy alta fiabilidad con una distancia máxima de 5 metros y una precisión de azimut de 1° dentro de una angulación limitada a 64°. El principio de medición está patentado y se basa en el momento de la medición de vuelo de la luz láser de pulso modulada Time-of-Flight (ToF). La principal ventaja de este sensor además de la precisión es que la velocidad de procesamiento es muy baja reduciendo el tiempo de cálculo.

## 2.2 Sistema de audio

La información que recibe el sistema de audio es la distancia del objeto y el ángulo con respecto a un plano que abarca desde el centro del sistema sensor hasta 32° hacia la izquierda del usuario y 32° a la derecha del usuario a una distancia máxima de 5 metros. Todo ello comprendido en un único plano horizontal a la altura del sistema sensor en que está integrado en las gafas en posición central.

El interfaz de audio da una respuesta al usuario en forma de clic sonoro procesados para este sistema a través de los auriculares el sonido es un clic de 2.048 muestras en frecuencia de muestreo de 44.100 Hz y de 47 ms. El sonido está convolucionado mediante Head Related Transfer Functions (HRTF) generando 64 HRTFs en acimut en cada 0,96° y 16 niveles en distancias de 0,5 metros hasta 5 metros.

Para las pruebas previas se ha utilizado Maximum Length Binary Sequence (MLBS) como fuente de sonido con el fin de medir las HRTF en una cámara anecoica.

La medición de HRTF se basa en el cálculo de la respuesta de impulso para ambos oídos mediante el filtro en el dominio de frecuencia. Para una señal de sonido  $x_1(n)$  reproducida por el altavoz, la respuesta registrada se puede calcular como:

$$Y_1 = X_1 LMF \quad (1)$$

donde  $X_1$  es la representación del sonido  $x_1(n)$  en el dominio de la frecuencia transformado,  $L$  es la función de transferencia del altavoz y todo el equipo de reproducción,



Figura 3. imágenes del dispositivo en uso.

F es la función de transferencia del espacio entre el altavoz y el oído humano y M es la transferencia función del micrófono del oído humano y todo el equipo de grabación.

Con el fin de calcular el filtro era necesario para generar un  $x_2(n)$  reproducido por los auriculares tales como la respuesta registrada Y2 para que sea igual a Y1. Y2 se puede calcular:

$$Y_3 = X_2HM \quad (2)$$

donde H es la función de transferencia de los auriculares y todos los equipos de reproducción.

Entonces filtro de T se obtiene a partir de (4) y (5).

$$T = \frac{LF}{H} \quad (3)$$

De esta manera el filtro de T se mide por las respuestas de impulso para cada oído.

La respuesta al impulso se obtiene mediante circular de correlación cruzada entre las MLBS de la entrada del sistema y la respuesta en la salida.

$$h(n) = \Omega_{sy}(n) = s(n) \Phi y(n) = \frac{1}{L+1} \sum_{k=0}^{L-1} s(k) y(n+k) \quad (4)$$

donde  $\Phi$  representar la correlación cruzada circular.

Transformada rápida de Hadamard (FHT) se utilizó para reducir el tiempo de cálculo para la generación de HRTF:

$$h(n) = \frac{1}{(L+1)s[0]} P_2(S_2\{H_{L+1}[S_1(P_1y(n))]\}) \quad (5)$$

donde P son las matrices de permutación, S - las matrices redimention e y HL + 1 es las matrices Hadamard de L + 1 grado.

En otras palabras, las señales binaurales influyeron en la generación de los sonidos virtuales. El desplazamiento de acimut de la fuente de sonido virtual se obtiene por el desplazamiento de fase interaural de las señales izquierda y derecha en términos de diferencia de tiempo entre el oído izquierdo y derecho. Cada sonido virtual corresponde a un pixel del módulo sensor y tiene una superficie de 64°.

### 3 Descripción de la prueba

Una vez verificada la fiabilidad del sistema y de los algoritmos se procedió a la realización de pruebas con personas invidentes en laboratorio. La prueba consiste en un ejercicio de detección de obstáculos, se creó un laberinto en el laboratorio mediante unas columnas formadas por cajas de cartón blando, con el objetivo de preservar de un posible impacto las personas invidentes

y que no se produjera ningún daño, la columna se desahacía al mínimo impacto.

El laberinto consistía en un pasillo de 14 m con 4 pares de columnas situados en paralelo a una distancia de 2,5m (Figura 4) en la que la persona invidente debía evitar los obstáculos y encontrar el pasillo libre [8].

El ejercicio consistía en los usuario del sistema debían mover la cabeza a izquierda y a derecha para que el sistema les indicara donde está situado el obstáculo y así encontrar el pasillo libre.

La localización para las pruebas de navegación en interior se correspondía con un gran salón con 15 m de longitud y una anchura de 10 m. El intervalo de tiempo entre cada conjunto de sonidos virtuales era 153 ms.

Las pruebas fueron realizadas por 10 personas invidentes y tuvo dos etapas diferenciadas en la primera se realizaron pruebas de adaptación del usuario con el sistema de navegación, para que conociera las particularidades del sistema y controles como el control del nivel de audio para adaptarlo a sus necesidades cognitivas.

Una vez familiarizados con el sistema los usuarios invidentes procedieron a realizar la prueba dentro del laboratorio del laberinto. Las pruebas se realizaron de dos formas la primera utilizando el bastón con el que están acostumbrados a navegar las personas invidentes y una segunda sin bastón.

## 4 Resultados

Lo primero que se realizó con los usuarios fue unas pruebas de adaptación al sistema para que fueran capaces de familiarizarse, todos fueron capaces en pocos segundos de percibir el origen del sonido y localizar los objetos que se probaron a modo de test. El tiempo medio de percepción del sonido era 2,32 minutos y para la detección de objetos de 4,86 minutos. El promedio de

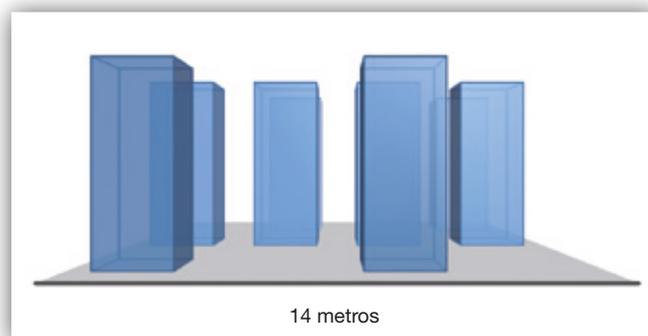
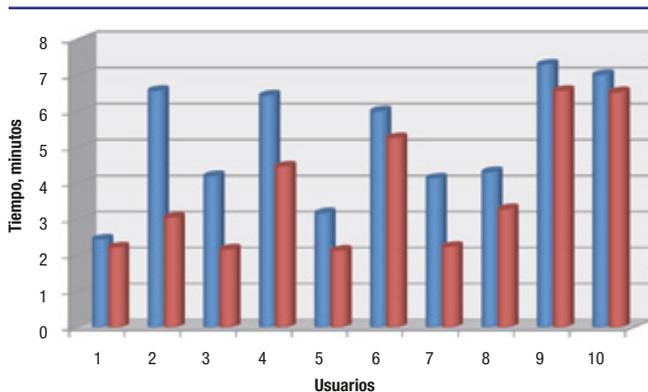


Figura 4. Prueba de Laberinto.



**Gráfico 1.** Tiempo de recorrido de las dos pruebas de laberinto.

tiempo total para completar estos ejercicios fue 3,59 minutos.

Una vez familiarizados con el sistema se realizó las pruebas en laberinto. Dentro del laboratorio se tomaron datos referentes al tiempo de recorrido de las 10 personas invidentes, se realizaron dos recorridos al laberinto por usuario, obteniéndose una mejora en el resultado con respecto al tiempo de superación de la prueba de laberinto como puede observarse en el Gráfico 1.

Los resultados sugieren que con el sistema utilizado es posible percibir la presencia, la posición y las dimensiones del objeto detectado. Esto es lo que indica que los sonidos virtuales son una solución prometedora como una parte de la interfaz de usuario de un sistema de navegación ciego. Los sonidos acústicos virtuales tienen la ventaja adicional de que necesitan un tiempo mucho más corto que el habla convencional.

Los datos actuales aumentan y amplían los estudios anteriores, que han demostrado la utilidad de las señales perceptuales directas a la navegación. En muy importante para el trabajo futuro para poder comparar los modelos de orientación en entornos más complejos, tales como estaciones de tren, supermercados, pasos elevados, etc... Otra línea de investigación es la localización de la fuente de sonido y señales sonoras características.

Los resultados actuales pueden tener buenas aplicaciones para el guiado en la navegación para las personas ciegas. Una ventaja del sistema es que podría integrarse con otros sistemas de navegación como el GPS y otros interfaces visuales.

## 5 Conclusiones

El sistema desarrollado ha permitido realizar las pruebas necesarias para demostrar que las personas invidentes

son capaces de localizar sonidos virtuales a través de auriculares.

También hay que destacar que han sido capaces de caminar sin el bastón blanco, haciendo uso sólo de las indicaciones que le facilita el sistema teniendo más confianza en el, a medida que aumentaba el tiempo de uso. Todos los usuarios solicitaban poder llevarse el dispositivo para hacer un uso diario del sistema. El dispositivo ha demostrado que es capaz de localizar los objetos de forma unívoca y de interpretarlo de forma audible para que una persona invidente sea capaz de localizar el objeto que tiene cercano y que no es capaz de detectarlo con un bastón. El experimento también demostró que las personas invidentes tienen muy desarrollada la capacidad auditiva y una gran capacidad de adaptación y aprendizaje en el entorno sonoro.

## Referencias

- [1] Blauert J. *Spatial Hearing: The Psychophysics of Human Sound Localization. Revised edn, The MIT Press, Cambridge, MA, USA, 1997.*
- [2] Brungart D.S.; Rabinowitz W. M. Auditory localization of nearby sources. Head-related transfer functions, *J. Acoust. Soc. Am.* 106(3), 1465-1479, 1999.
- [3] Hartmann B. M. Localization of the sounds in rooms. *J. Acoust. Soc. Am.* 74, pp. 1380-1391.
- [4] Andrew Brughera, Larisa Dunai, and William M. Hartmann. Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.* Volume 133, Issue 5, pp. 2839-2855 (2013)
- [5] Fitzpatrick D. C., Kuwada S., and Batra R. Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model. *The Journal of Neuroscience*, February 15, 2000, 20(4):1605-1615.
- [6] Takahashi T. T.; Keller C. H. Representation of multiple sound sources in the owl's auditory space map. *The Journal of Neuroscience*, 14(8), pp. 4780-4793, 1994.
- [7] Mora J.L. G.; Rodríguez-Hernández A.F.; Martín F., Castellano M.A. Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people. *Proceedings of the 2<sup>nd</sup> Information and Communication Technologies Conference, ICTTA'06, IEEE* pp. 837-842, 2006.
- [8] Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnes, Beatriz Defez. Dispositivo de navegación para personas invidentes basado en la tecnología time of flight. *Dyna*, Edición 179, pp. 33-41, Medellín, junio, 2013.



# Audiotec

Ingeniería y Control del Ruido

Mapas de Ruido. Planes de Acción  
Ensayos y Mediciones Certificadas  
Aislamiento y Acondicionamiento Acústico  
Certificación de Productos en Cámaras Normalizadas  
Proyectos de Soluciones Acústicas: industria y edificación  
Centro de Formación y Consultoría



Centro Tecnológico de Acústica

902 37 37 99

[www.audiotec.es](http://www.audiotec.es)

[info@audiotec.es](mailto:info@audiotec.es)

Madrid, Barcelona, Valladolid, Valencia, Bilbao, Cantabria, La Rioja y Zaragoza



empresa  
familiarmente  
responsable



[www.audiotec.es](http://www.audiotec.es)

