

# El sonido tridimensional en la realidad virtual

Julio Manuel Vega Pérez

*Universidad Rey Juan Carlos, Móstoles (Madrid), Spain*

1 de junio de 2009

## 1. Resumen

*En este trabajo se pretende dar un vistazo amplio al tratamiento e influencia de los sonidos 3D en diversas aplicaciones como videojuegos y, en general, escenarios 3D destinados a causar una inmersión en él con sensaciones cercanas a la realidad. En este aspecto es de destacar el sonido virtual en 3D porque, si nos fijamos un poco en nuestro sistema de audición, estamos constantemente analizando sonidos y gracias a él somos capaces de localizar objetos en el espacio que nos rodea, sin utilizar ningún otro sentido.*

## 2. Introducción

Un ejemplo bastante ilustrativo del sistema de audición humano lo encontramos en el famoso libro de Albert Bregman *Auditory Scene Analysis*, donde describe la siguiente analogía: supongamos que estamos frente a un lago, del cual salen dos surcos largos y angostos hacia la orilla. En cada surco se coloca una hoja de árbol, la cual se moverá a medida que el agua del lago fluya hacia los surcos. Sólo analizando el movimiento de ambas hojas, sin mirar nada más que las hojas, la idea es poder responder preguntas como:

1. ¿Cuántos barcos hay en el lago y dónde están?
2. ¿Qué barco está más cerca?

3. ¿Cuál es más potente?
4. ¿Hay viento?
5. ¿Ha caído algún objeto grande sobre el lago?
6. ¿Hay algún nadador y cómo nada de rápido?

Esto puede parecer una misión imposible. Pero la verdad es que el sistema auditivo humano hace este tipo de análisis todo el tiempo y a una velocidad extremadamente alta. Nuestros tímpanos equivalen a las hojas depositadas en los surcos. Nuestro sistema auditivo, basándose únicamente en el análisis del patrón de variaciones en ambos tímpanos, es capaz de generar una idea muy precisa de nuestro entorno auditivo.

Suponiendo que nos encontremos en una fiesta muy ruidosa, lo que típicamente se conoce en inglés como el *cocktail party problem*, donde hay muchas fuentes sonoras de todo tipo interactuando al mismo tiempo, no necesitamos realizar ningún esfuerzo para responder acertadamente a preguntas como ¿Cuánta gente está hablando? o ¿Quién suena más fuerte o está más cerca? o ¿Hay alguna máquina haciendo ruido de fondo?. El sistema auditivo humano se extiende desde el oído externo hasta la corteza cerebral y es capaz de realizar una cantidad impresionante de operaciones complejas en muy poco tiempo.

Entonces ¿qué es un sonido en 3D? Es un sonido que trata de crear una imagen mental acústica de

las fuentes de sonido espacial dentro de un espacio artificial.

### 3. Aplicaciones

#### 3.1. Mediazine

Pues bien, Los sistemas de sonido tridimensional o la comunicación entre el ser humano y la máquina son algunos de los campos de trabajo del Instituto Fraunhofer de Tecnología de Medios Digitales (IDMT). A grandes rasgos podría decirse que en las nuevas instalaciones de Ilmenau, operativas desde 2008, lo que se busca es aproximar los mundos virtual y real.

El director de este instituto fue el creador del algoritmo del formato de audio que más éxito tiene actualmente a nivel mundial: el MP3. El sistema de codificación de audio MP3 revolucionó el sector de la música y la forma de oír música. La posibilidad de comprimir los datos audio a la doceava parte de su tamaño original abrió paso a los reproductores MP3 y a las descargas de música por Internet.

Un sistema bastante interesante desarrollado por este instituto es el descrito en [Schiffner y Chodura, 2000], llamado Mediazine. Se trata de un magazine multimedia interactivo, y en un entorno distribuido como es Internet. Incluye, entre otras cosas, un servidor de audio espacial que maneja las comunicaciones por teléfono, o aplicaciones como la televisión y radio. Y es espacial porque proporciona una salida binaural de modo que, por ejemplo, en el entorno virtual los participantes pueden situar mentalmente al resto de avatares, sólo por el sonido.



Figura 1: Habitación virtual de Mediazine

#### 3.2. Asistencia en viaje

Según estudios, muchos de los accidentes que se producen en carretera se podrían evitar si el tiempo entre la detección del peligro y la reacción al respecto disminuyera. De modo que cuanto más rápido es la reacción, mayor es la probabilidad de evitar el peligro.

Actualmente, el sistema visual es el sentido más utilizado, ya que reporta mucha información al conductor de forma rápida. Sin embargo, debido a la incremental aparición de controles en un coche, la visión acaba por sobrecargarse. Pues bien, en el trabajo de [Bellotti *et al.*, 2000] se presenta un sistema que se centra en la capacidad del sistema auditivo para reforzar las habilidades de un conductor. En ese estudio se analizan tres posibles casos de control auditivo: *(i) estado lateral*, que ayude al conductor a mantenerse por el carril; *(ii) estado longitudinal*, para regular la distancia de seguridad con los coches que lo rodean; y *(iii) situación de aparcamiento*, para ayudar en maniobras de estacionamiento del vehículo.

El sonido en 3D permite la reproducción del sonido desde cualquier posición y en cualquier trayectoria. A parte también se tratan aspectos como la intensidad o el contenido espectral, de modo

que la percepción del sonido sea óptima.

Un problema a tener en cuenta para el diseño del sistema es el conocido como *Efecto Franssen*. Si consideramos dos fuentes de sonido en diferentes posiciones, y una emite primero que la otra el mismo tono. Aunque la primera fuente de sonido pare de sonar, y continúe la segunda, los estudios revelan que las personas seguimos escuchando el sonido desde la posición que emitía primero y que actualmente está parada.

### 3.3. Simulación de conducción

En la misma línea que el anterior trabajo comentado, vamos a citar otro estudio al respecto, [Xun-xiang *et al.*, 2008], que trata de añadir el efecto del sonido 3D a la inmersión en un entorno de conducción de automóviles, principalmente porque el campo de visión en estos sistemas virtuales suele ser limitado, y para satisfacer tales limitaciones.

Los aspectos a tener en cuenta para una simulación auditiva correcta son: (i) *principio de posicionamiento del audio espacial*, que nos puede reportar información acerca de la posición de las fuentes de sonido o velocidad de movimiento (en su caso); (ii) *modelo de distancia*, que estudia cómo decae la intensidad de la voz hasta que se pierde; y (iii) *efecto Doppler*, cuya consecuencia radica en la sensación de percibir un sonido con frecuencia cambiante (inexistente), debido al relativo movimiento entre el observador y la fuente de sonido.

#### 3.3.1. OpenAL

Es una plataforma de desarrollo de audio 3D. Similar a *OpenGL*, con esta librería podemos programar a alto nivel; además tampoco hace un uso excesivo de CPU, sino que su procesamiento recae

mayormente en la GPU. Nos ofrece una serie de efectos sonoros, como la reflexión, reverberación, oclusión, obstrucción, o el efecto Doppler.

En un entorno de *OpenAL*, tan sólo hay un observador, que representa a una persona con dos oídos. Asimismo tenemos un buffer donde cargamos los archivos de sonido, a los cuales asignamos una determinada posición en el entorno virtual; o varias posiciones, de forma que sólo tenemos que cargar cada sonido una vez aunque lo reproduzcamos desde varias zonas.

### 3.4. Ayuda a personas con discapacidad

Los entornos virtuales basados en el audio 3D también se suelen usar como medio para motivar y mejorar la atención de personas con ciertas discapacidades. En el proyecto *AudioChile*, desarrollado por la *Universidad de Chile*, éste es el principal propósito.

Un trabajo interesante lo tenemos en [Sánchez y Sáenz, 2005], donde se centran en usar el sonido 3D como interfaz para mejorar las habilidades de los usuarios en el reconocimiento de espacios virtuales a través del sonido. Los niños con discapacidades visuales tienen una percepción distinta del mundo que les rodea, frente a los niños que ven con normalidad. De modo que suelen desarrollar otros sentidos, con el oído, que les ayuden a orientarse y, por tanto, a moverse por el espacio. Así, necesitan aprender a interpretar esos sonidos.

En *AudioChile*, se proporcionan una serie de escenarios virtuales que recrean las diferentes regiones de Chile. Los niños tienen que navegar por estos entornos interactuando a través del sonido y la música, teniendo que realizar determinadas actividades así como tomar ciertas decisiones dentro de tales mundos.

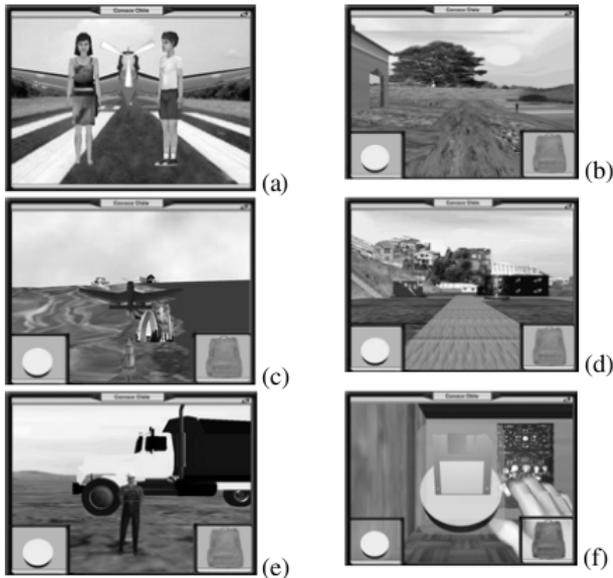


Figura 2: Distintas situaciones del entorno virtual *AudioChile*

Otro trabajo también desarrollado por la *Universidad de Chile*, pero que se centra en los niños ciegos, lo encontramos en [Lumbreras y Sánchez, 1999]. Hemos visto cómo el sonido 3D puede ayudar a interactuar en entornos 3D, cuando la visión del mismo es limitada; sin embargo esta interacción se torna más difícil si el usuario no tiene ninguna visión.

En este estudio se demuestra que a través del audio 3D los usuarios se pueden crear una imagen mental del entorno virtual por el cual están navegando. Normalmente la visión juega un papel fundamental en nuestro conocimiento del mundo; la mayoría de las representaciones mentales se hacen a través de la percepción visual.

Pues bien, la aplicación aquí desarrollada se denomina *AudioDoom*, y está formada por una serie de escenarios virtuales, un avatar manipulado por el usuario, un conjunto de objetos dinámicos y una serie de interacciones entre esas entidades que constituyen el argumento de la historia.

Para interactuar en *AudioDoom*, los niños operan a través del espacio de sonido virtual, actuando

sobre *voxels*. Éstos son una discretización del espacio, simplificando las posiciones de interacción y creando un repositorio concreto para una determinada entidad. Tal entidad normalmente es un objeto virtual representado acústicamente (una puerta, una caja, un personaje, etc.).

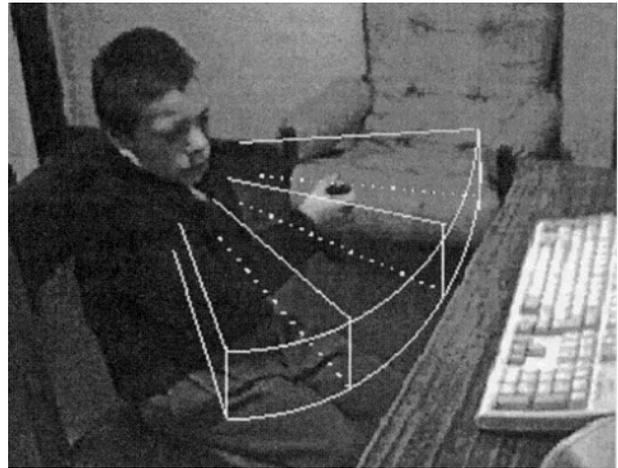


Figura 3: El niño explora *AudioDoom* interactuando con entidades situadas en *voxels* virtuales (dibujadas artificialmente sobre la fotografía)

## 4. Problemas

Para renderizar el sonido 3D, hay que tener en cuenta muchos aspectos, sobretudo en aplicaciones en tiempo real. Al igual que ocurre con los gráficos, uno de los factores que más influye es el número de polígonos empleados en una escena. En el caso que nos estamos centrando, concretamente son los polígonos de reflexión del sonido. Además, dado que las tarjetas gráficas cada vez ofrecen un mejor rendimiento, el número de éstos hoy en día suele ser muy grande.

El sonido 3D virtual requiere que se computen las siguientes cualidades:

- *Trasmisión*. La absorción del sonido por el aire, distribución de energía, y también cómo afectan las diferentes superficies en el mismo.

- *Reflexión*. Propagación del sonido por el entorno.
- *Reverberación*. Ligera prolongación del sonido una vez que se ha extinguido el original, debido a las ondas reflejadas.
- *Función de transferencia relativa a la cabeza (HRTF)*. Cambios de propagación/interferencia del sonido debido a la distribución de las orejas en la cabeza. Efecto tratado más detalladamente en el trabajo de [Wu *et al.*, 1996].
- *Difracción*. Dispersión y curvado aparente de las ondas cuando encuentran un obstáculo.
- *Refracción*. Cambio de dirección que experimenta una onda al pasar de un medio material a otro.

Existen numerosos trabajos concernientes a cómo optimizar las escenas virtuales para conseguir un sonido 3D apto para tiempo real. Nosotros nos centraremos en algunos de ellos. Para comenzar, en el trabajo de [Joslin y Magnenat-Thalmann, 2003] se centran en reducir el número de polígonos de la escena. La motivación es sencilla; mientras que para gráficos es necesario un gran número de polígonos para tener gran calidad de imagen, en audio no es necesario tener tal número de polígonos, debido a que la longitud de onda del sonido es muy grande (17m para una baja frecuencia de 20Hz). De modo que con una representación sencilla, podemos obtener un sonido renderizado con bastante precisión.

Así, en el mecanismo empleado en tal estudio, se quedan con las facetas más significativas de escenas complejas. Para ello, emplean segmentación de la escena, *clustering* de puntos 3D y agrupamiento en cajas contenedoras.

En el trabajo de [Tsingos *et al.*, 2004] se centran en el problema de tener varias fuentes de sonido móviles. En este caso, la idea es emplear mecanismos de *culling* y, en definitiva, emplear el nivel de detalle de la escena adecuado para poder tener varias fuentes de sonido simultáneamente y en tiempo real.

El procedimiento seguido es eliminar las fuentes de sonido inaudibles en un determinado momento, y agrupar las restantes en clusters. El *pipeline* seguido es el que se muestra en la imagen 4.

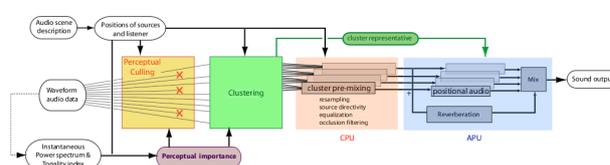


Figura 4: *Pipeline* usado, empleando *culling/clustering* para reducir el número de fuentes y amortizar los costes de computación sobre grupos de sonidos

La misma ideología se puede apreciar en otro trabajo del mismo grupo de investigación ([Moeck *et al.*, 2006]). La novedad aquí reside en un *clustering* jerárquico y en una mezcla conveniente de sonidos para el tratamiento de éstos de forma conjunta. El pipeline es el de la figura 5.

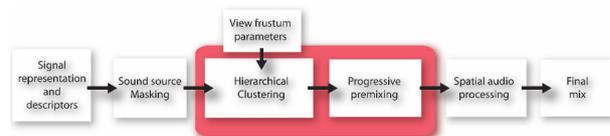


Figura 5: *Pipeline* usado, empleando *clustering* jerárquico

Otros trabajos cuyo interés está centrado en reducir los costes computacionales en el tratamiento de sonidos 3D, se pueden consultar en la bibliografía. Véase por ejemplo [Doerr *et al.*, 2007] o [Kim *et al.*, 1997] entre otros...

## Referencias

- [Bellotti *et al.*, 2000] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, y Massimiliano Margarone. Using 3d sound to improve the effectiveness of the advanced driver assistance systems. *Department of Electronics and Biophysical Engineering, University of Genova, Italy*, 2000.
- [Cowan y Kapralos, 2008] Brent Cowan y Bill Kapralos. Spatial sound for video games and virtual environments utilizing real-time gpu-based convolution. *Faculty of Business and Information Technology, Health Education Technology Research Unit, University of Ontario Institute of Technology, Oshawa, Ontario, Canada*, 2008.
- [Dobler y Stampfl, 2004] Daniel Dobler y Philipp Stampfl. Enhancing three-dimensional vision with three-dimensional sound. *SIGGRAPH 2004 Course, Klimschgasse, Viena, Austria*, 2004.
- [Doerr *et al.*, 2007] Kai-Uwe Doerr, Holger Rademacher, Silke Huesgen, y Wolfgang Kubbat. Evaluation of a low-cost 3d sound system for immersive virtual reality training systems. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 2, MARCH/APRIL 2007*, 2007.
- [Gaye, 2002] Lalya Gaye. A flexible 3d sound system for interactive applications. *Future Applications Lab, Viktoria Institute, Sweden*, 2002.
- [Heuten *et al.*, 2006] Wilko Heuten, Daniel Wichmann, y Susanne Boll. Interactive 3d sonification for the exploration of city maps. *University of Oldenburg, Oldenburg, Germany*, 2006.
- [Joslin y Magnenat-Thalmann, 2003] Chris Joslin y Nadia Magnenat-Thalmann. Significant facet retrieval for real-time 3d sound rendering in complex virtual environments. *MIRALab, University of Geneva, Switzerland*, 2003.
- [Kim *et al.*, 1997] Tae-Sun Kim, Seh-Woong Jeong, Byung-Chul Park, y Sang-II Park. New real-time implementation of 3d-sound system using tla algorithm. *IEEE Transactions on Consumer Electronics, Vol. 43, No. 3, AUGUST 1997*, 1997.
- [Lumbreras y Sánchez, 1999] Mauricio Lumbreras y Jaime Sánchez. Interactive 3d sound hyperstories for blind children. *Department of Computer Science, University of Chile, Chile*, 1999.
- [Matsumura *et al.*, 2005] Tomoya Matsumura, Nobuyuki Iwanaga, Wataru Kobayashi, Takao Onoye, y Isao Shirakawa. Embedded 3d sound movement system based on feature extraction of head-related transfer function. *IEEE Transactions on Consumer Electronics, Vol. 51, No. 1, FEBRUARY 2005*, 2005.
- [Moeck *et al.*, 2006] Thomas Moeck, Nicolas Bonneel, Nicolas Tsingos, George Drettakis, Isabelle Viaud-Delmon, y David Alloza. Progressive perceptual audio rendering of complex scenes. *REVES/INRIA, Sophia-Antipolis, France*, 2006.
- [Schiffner y Chodura, 2000] Norbert Schiffner y Hartmut Chodura. Mediazine - a combination of television, radio, www, telecommunication and 3d computer sound and graphics. *Fraunhofer Institute for Computer Graphics, Fraunhofer, Germany*, 2000.

- [Sánchez y Sáenz, 2005] Jaime Sánchez y Mauricio Sáenz. 3d sound interactive environments for problem solving. *Department of Computer Science, University of Chile, Chile, 2005.*
- [Tsingos *et al.*, 2004] Nicolas Tsingos, Emmanuel Gallo, y George Drettakis. Perceptual audio rendering of complex virtual environments. *REVES/INRIA, Sophia-Antipolis, France, 2004.*
- [Wu *et al.*, 1996] Jiann-Rong Wu, Cha-Dong Duh, y Ming Ouhyoung. Head motion and latency compensation on localization of 3d sound in virtual reality. *Communication and Multimedia Lab., Department of Computer Science and Information Engineering, National Taiwan University, Taiwan, 1996.*
- [Xun-xiang *et al.*, 2008] Li Xun-xiang, Zhan Duo, Cheng Dongyan, y Chen Dingfang. Implementation of 3d sound effect modeling technology based on vehicle driving simulator. *Modern Industrial Design Institution, College of Computer Science, Zhejiang University, Hangzhou, China, 2008.*