

Sistemas de auralización y sonido 3D para su aplicación en entornos virtuales de edificios del patrimonio histórico-arquitectónico

REFERENCIA PACS: 43.55.-Ka, 43.60.-c

Montell, Radha¹; Segura, Jaume²; Vera, Lucía²; Barba, Arturo¹; Giménez, Alicia¹; Fernández, Marcos²; Cibrián, Rosa³; Cerdá, Salvador⁴; Romero, José⁵

1 E.T.S.I.Industriales., Univ. Politécnica de Valencia, Camí de Vera s/n, radmonse@fiv.upv.es, arbase@doctor.upv.es, agimenez@fis.upv.es

2 Institut de Robòtica, Universitat de València, Poligon de la Coma s/n, jsegura@uv.es, vemalu@uv.es, mfmartin@uv.es

3 Facultad de Medicina. Universidad de Valencia, rosa.m.cibrian@uv.es

4 E.T.S.I.Arquitectura, Univ. Politécnica de Valencia, salcerjo@mat.upv.es

5 Escuela Politécnica Superior de Gandía, Universidad Politécnica de Valencia, romerof@fis.upv.es

Abstract

This paper presents the work that is being developed within the Project BIA2008-05485 related to 3D sound applications in virtual environments for the representation and/or refurbishment of buildings in the Historic Patrimonial Catalogue.

The application of auralization techniques and of sound rendering will allow us to evaluate different proposals of the refurbishment in churches, monasteries, etc., which nowadays have several uses.

Resumen

En este artículo, se presenta el trabajo que se está desarrollando en el proyecto BIA2008-05485 en relación con la aplicación de sonido 3D a entornos virtuales para la representación y/o rehabilitación de edificios del patrimonio histórico-arquitectónico.

La aplicación de técnicas de auralización y de presentación sonora permitirá la evaluación subjetiva de propuestas de rehabilitación de iglesias, monasterios, etc., que actualmente realizan múltiples funciones.

1. INTRODUCCIÓN

La aplicación de sistemas de sonido 3D a entornos virtuales y de simulación ha dado resultados muy interesantes en campos como la industria de los video-juegos, los simuladores y demostradores de conducción y manejo de vehículos ligeros y pesados, etc. [1][2][3]

Desde el punto de vista del medio, la combinación de técnicas de simulación acústica ha permitido la optimización de la predicción de parámetros acústicos de una sala como paso previo a su auralización. Según M. Kleiner, la auralización es un proceso de presentación audible, por modelización física o matemática del campo sonoro de una fuente en un espacio, de tal forma que simulan la experiencia sonora en una determinada posición en el espacio modelado. Kleiner presenta la auralización del campo sonoro 3D de forma análoga a la

visualización en entornos virtuales.[4]. Entre las técnicas de auralización más conocidas encontramos:

- Auralización completamente calculada: a partir del modelo de la sala, determina la respuesta impulsiva de la sala (RIR) y de ella su correspondiente binaural (BRIR), a partir de ella y mediante procesado de la señal ésta puede ser convolucionada con ficheros de audio anecoicos.
- Auralización multifuente calculada: a partir de la RIR multicanal se puede realizar la convolución de diversos canales para un sistema multifuente.
- Auralización directa a partir de un modelo a escala: mediante el modelo físico a escala y usando técnicas ultrasónicas (realizan un escalado directo permitiendo una convolución en tiempo real).
- Auralización indirecta a partir de un modelo a escala: con el modelo físico a escala y usando técnicas ultrasónicas (miden la respuesta impulsiva binaural de la sala).

Además, la presentación de la señal acústica en entorno virtual puede realizarse mediante técnicas de procesado de señal y reproducción sonora. Entre las técnicas más conocidas de representación sonora encontramos: técnicas binaurales mediante el uso de los HRTF, técnicas estéreo con cancelación cross-talk, técnicas de 'panning', otras técnicas multicanal (ambisonics, WFS, multicanal 5.1, etc). Estas técnicas permiten entre otras aplicaciones, la localización del sonido en salas.[5][6]

En este artículo se plantea el estudio de un protocolo de evaluación mediante la combinación de simulación acústica y técnicas de realidad virtual para la determinación de los parámetros acústicos en edificios de nuestro entorno socio-cultural. Este trabajo se engloba en el proyecto financiado por el Ministerio de Ciencia e Innovación con referencia BIA2008-05485. El proyecto pretende proporcionar herramientas para la valoración acústica objetiva y subjetiva de una sala antes de su construcción o remodelación, mediante el aporte de una metodología concreta para analizar y obtener la calidad necesaria en edificios donde la audición sea un factor importante de diseño. Para ello se parte de la hipótesis que si se diseña adecuadamente un entorno de realidad virtual y se realiza en él una auralización, es posible evaluar la calidad de la sala en todos sus puntos antes de realizar la rehabilitación de la sala o su propia construcción.

2. METODOLOGÍA

En el proceso de auralización se diferencian cuatro etapas: (1) Modelización, (2) Cálculo de la respuesta de la sala; (3) Procesamiento digital de la señal y (4) Presentación de la señal procesada. En la figura 1 se puede ver este proceso incluido dentro de la estructura del sistema global.

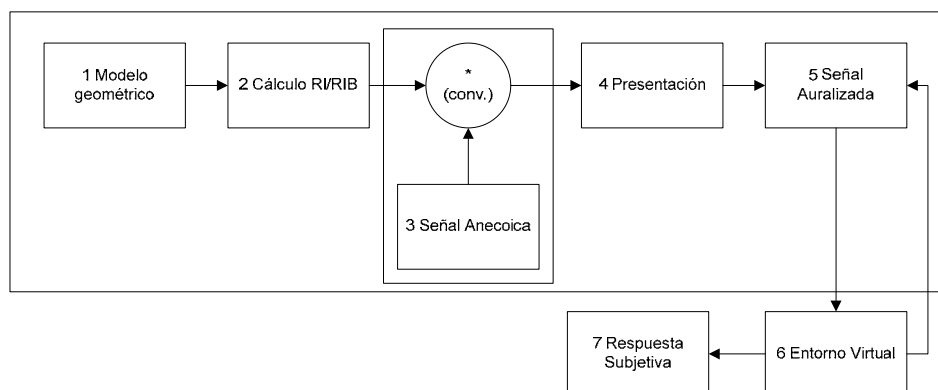


Figura 1: Esquema del proceso de auralización integrado en nuestro sistema.

2.1. Proceso de Auralización

En la primera fase, se modelan las propiedades del espacio acústico (geometría, parámetros de las superficies, fuente(s), receptor(es), etc.) y se aplica un algoritmo de simulación de las propiedades de transmisión (trazado de rayos, modelo de fuentes espejo, métodos de elementos finitos, métodos híbridos, etc). En esta etapa se obtiene lo que se conoce como respuesta impulsiva (RI). Ésta se subdivide en tres partes: sonido directo, reflexiones tempranas y reverberación.

En la segunda fase se obtiene la Respuesta Impulsiva Binaural (RIB). Ésta representa la respuesta de la sala para una fuente y un receptor particular. Las características del receptor son incorporadas mediante el uso de la Función de Transferencia Relacionada con la Cabeza (HRTF). Mediante estos filtros se modela el efecto de las reflexiones sobre la cabeza, el torso y el pabellón auricular. Al igual que las características anatómicas son propias de cada individuo, los filtros HRTF también lo son. Éstos se pueden medir del oyente, posicionando unos pequeños filtros en la entrada del canal auditivo, o de maniquís diseñados para ello. En los casos en los que no es posible obtener HRTF individualizados, existen bancos de filtros ya creados [7]. La señal binaural se sintetiza mediante la convolución de los HRTF correspondientes a cada oído (HRTF_{izq.} - HRTF_{der.}) sobre la RI monoaural. La utilización de HRTFs dependerá del método de presentación que se quiera utilizar.

En la tercera fase, el objetivo es obtener la señal de audio espacializada de acuerdo con el espacio modelado. Para esto, la señal acústica de la fuente (usualmente grabada en cámara anecoica) es convolucionada con la Respuesta Impulsiva Binaural obtenida en la fase anterior.

La cuarta fase consiste en reproducir la señal en los oídos del receptor de forma que la impresión perceptual sea equivalente a la que se obtendría en el espacio original. Los métodos de presentación se dividen en aquellos que reproducen la señal acústica mediante auriculares (técnicas binaurales) y los que la reproducen mediante altavoces (técnicas transaurales) [8][9][10]. Los HRTF antes nombrados son utilizados en las técnicas binaurales. Una vez obtenida la señal para cada oído, ésta llega a él a través de los auriculares. A pesar de ser una técnica ampliamente utilizada, el hecho de tener los oídos tapados produce una serie de efectos no deseados, como por ejemplo, la localización de fuentes dentro de la cabeza. No obstante, existen técnicas para reducir estos efectos [11].

Las técnicas binaurales también son utilizadas con estructuras de altavoces (normalmente dos) y simulan el efecto de los auriculares. En este caso se les denomina técnicas transaurales [12][13][14]. En éstas, es necesaria la implementación de filtros de reducción de la diafonía (*Cross-talk cancellation*, en inglés) para que las señales emitidas no se interfieran y que cada oído perciba la señal que le corresponde. Éstas técnicas tienen un uso limitado debido a que solo existe una zona concreta, llamada *sweet spot*, donde el sonido es correctamente renderizado. Además presentan problemas de reversión frontal-trasero donde el receptor localiza incorrectamente la procedencia de la señal. No obstante, este problema puede ser solucionado situando altavoces detrás del receptor (técnicas doble-transaurales). Dentro de las técnicas transaurales podemos encontrar:

- Técnicas de paneo [15][16]: Consisten en alimentar todos los altavoces con la misma señal monofónica, y reproducirla variando la amplitud (paneo en amplitud) o el desplazamiento en el tiempo (paneo temporal). Las técnicas de paneo en amplitud son utilizadas para crear fuentes “fantasma” posicionadas virtualmente entre altavoces adyacentes. Esta técnica se aplica tanto en 2D como en 3D, creando estructuras de altavoces en línea o en triángulos, respectivamente. Existen varias técnicas de paneo en amplitud 2D según el número de altavoces: estereofonía (2 altavoces), cuadrafonía (4 altavoces) o los sistemas 5.1, 6.1 y 7.1, utilizadas en las instalaciones de Home Cinema. En la actualidad existen sistemas de grabación multicanal que combinan la entrada de varios micrófonos para crear sonidos multicanal. Por otra parte, dentro de las técnicas de paneo en amplitud en 3D se encuentra el *Vector Base Amplitude Panning* (VBAP) [15]. Respecto a las técnicas de paneo temporal, su principal uso es crear efectos sonoros, como por ejemplo, simular el movimiento de la fuente hacia un altavoz emitiendo desde éste la señal en primer lugar.

- Ambisonics [17][18]: Este método se basa en la combinación de componentes armónicos esféricos para codificar y reproducir el sonido espacial. En el caso de armónicos de primer orden, se denomina *B-format*. En éste, la información es codificada en cuatro canales (uno para cada componente direccional, X, Y, Z y uno para la no-direccional, W) mediante el uso de un micrófono *Soundfield*. En cuanto a la decodificación, existen múltiples métodos con diferentes ecuaciones de decodificación según la aplicación que se quiera dar. Ésta técnica se puede extender a órdenes mayores [19].
- Síntesis de campo sonoro (*Wave Field Synthesis*)[20]: El objetivo de esta técnica es reproducir exactamente los frentes de onda dentro de una zona amplia de escucha, de forma que varios oyentes pueden moverse libremente por la zona manteniendo una correcta percepción de la localización de las fuentes virtuales. Este es un tipo de reproducción holofónica que, al igual que los hologramas en el contexto visual, permite conservar las características espaciales de distancia y dirección. Basándose en el principio de Huygens, los puntos donde se encuentran los micrófonos pueden verse como fuentes de sonido primarias, de forma que el campo de ondas original es reconstruido a partir de los frentes emitidos por las diferentes fuentes. En la práctica, esta técnica requiere de un gran número de altavoces dispuestos en línea recta y gran cantidad de recursos computacionales, ya que cada altavoz debe ser alimentado con la señal correspondiente, y esta disposición de altavoces solo permite la reproducción horizontal del campo de onda.

Cuando estas técnicas se utilizan en un entorno de realidad virtual, es posible aumentar el realismo mediante el uso de sistemas de seguimiento del movimiento de cabeza. Wenzel et al. [21] desarrolló el denominado "Convolutron", que permite seleccionar los filtros HRTF correspondientes según la posición relativa de la cabeza del receptor respecto de la fuente.

2.2. Representación de Escenas

Una escena es una representación de datos, con elementos audiovisuales en un ordenador. Normalmente se suelen considerar estos elementos como objetos generados en el entorno de un lenguaje de programación como C++ o Java. La creación de una escena a partir de estos objetos puede consistir en la agrupación o mezcla de éstos para formar otros objetos sonoros y gráficos de orden mayor. Las escenas se suelen definir a partir de librerías preprogramadas, llamadas APIs (*Application Programming Interface*), las cuales contienen clases, funciones y propiedades que permiten definir el comportamiento y el estado de los objetos de la escena. La descripción de una escena sonora en un entorno virtual suele requerir la definición de un modelo de reverberación, relacionado con la propia sala y sus propiedades, que describa también la respuesta al impulso de la sala y un modelo de receptor-fuente que describa la atenuación con la distancia, la directividad de fuente y receptor, etc.

Entre las APIs más conocidas y usadas para trabajar con sonido 3D encontramos:

- OpenAL: Interfase multi-plataforma de código libre. Soporta un contexto sonoro por tarjeta de sonido, un receptor por contexto y tantas fuentes como pueda soportar la computadora. Proporciona atenuación basada en la distancia, soporte para simular el efecto Doppler y soporta modificación de la frecuencia en tiempo de ejecución. [22]
- Microsoft Direct Sound 3D (DS3D): API desarrollada por Microsoft. Solo funciona sobre Windows. El modelo de atenuación que proporciona la API se basa en la distancia fuente-receptor y en la directividad de la fuente. Incluye una serie de efectos sonoros limitados pero fáciles de aplicar: eco, efecto coro, distorsiones, compresión, *gargle*, *flange*, ecualizador paramétrico, reverberación de ondas y un modelo de reverberación ambiental con las siguientes propiedades: atenuación, nivel de atenuación de la señal reflejada (*roll-off factor*), tiempo de retardo de las reflexiones respecto el sonido directo, tiempo de retardo de las reverberaciones respecto las primeras reflexiones, difusión y densidad. [23]
- Extensión EAX: Librería de efectos sonoros acelerada por hardware desarrollada por Creative. Se puede utilizar tanto sobre DS3D como sobre openAL. Posee un potente conjunto de efectos para simular el modelado del campo sonoro: oclusión, obstrucción, reverberación, posibilidad de renderizar varios entornos acústicos simultáneamente y, en la

última versión (5.0), se proporciona soporte para 128 sonidos simultáneos, permite el procesamiento en tiempo real de la señal de micrófono, permite controlar sobre qué altavoz se reproduce una señal determinada, permite la reproducción de hasta cuatro efectos de sonido simultáneos, permite el modelado de efectos muy cercanos al receptor y mejora el sistema de oclusión permitiendo que tanto el sonido directo como las reverberaciones sean afectadas. [24]

- Java3D: Es una extensión del lenguaje Java. Permite modelar reverberación, absorción del aire y proporciona una definición de la directividad dependiente de la frecuencia. [25]
- VRML/X3D: Esta interfaz permite modelar entornos virtuales de forma sencilla y representa las escenas mediante grafos de escena. Proporciona un modelo de atenuación con la distancia, definiendo los límites mediante dos elipses. [26]
X3D es el sucesor de VRML e incorpora nuevas características gráficas (mundos multi-usuario, animación de cuerpos, etc.). No obstante, no incorpora nuevas funcionalidades para el modelado del campo acústico. [27]
- MPEG-4 Binary Format for Scenes (BIFS): BIFS es un lenguaje de descripción de escenas. Proporciona un modelo de atenuación dependiente de la distancia y permite definir la absorción del aire y la velocidad del sonido en el medio. Permite modelar la reflectividad y la transmisión de las superficies, así como la reverberación (tiempo de reverberación dependiente de la frecuencia, retraso respecto el sonido directo y nivel de reverberación). Permite caracterizar la acústica de salas tanto desde el punto de vista físico como perceptual. [28]

3. APLICACIONES

Se ha modelado un entorno virtual en VRML 2.0 (figura 2), en el cual se han creado varios nodos sonido asociados a las posiciones donde se ha modelado gráficamente la forma de la fuente y se han fijado los parámetros correspondientes a las elipses que definen la atenuación. Se ha ajustado la orientación de la fuente y se ha activado el efecto de sonido 3D. En este entorno virtual el usuario puede interactuar pulsando sobre la representación de las fuentes.



Figura 2: Entorno virtual modelado en VRML. Modelización de instrumentos gráfica y acústica.

En el siguiente entorno virtual (figura 3), se ha modelado el Paraninfo de la Universidad Politécnica de Valencia, “sala piloto” en nuestro estudio, con la aplicación 3D Studio Max.



Figura 3: Paraninfo UPV modelado en 3D Studio Max

En ambos entornos virtuales, el campo acústico se simula mediante la reproducción de un fichero de sonido auralizado [29] para cada posición fuente-receptor, actualizado a través de la retroalimentación del sistema gráfico al acústico.

4. CONCLUSIONES

En este trabajo se ha planteado un sistema que combina la simulación gráfica y acústica para la creación de entornos virtuales. Se ha estudiado el proceso de auralización y algunas de las APIs más utilizadas en la creación de estos entornos.

Este estudio entra en el marco del proyecto referenciado que engloba recintos patrimoniales de la Comunidad Valenciana. Con ello pretendemos proporcionar una herramienta en los entornos real-virtual, para la evaluación acústica objetiva-subjetiva de un recinto antes de su construcción, o en su mejora en procesos de rehabilitación y acondicionamiento.

5. AGRADECIMIENTOS

Este trabajo ha sido subvencionado por el Ministerio de Ciencia e Innovación mediante el proyecto de investigación BIA2008-05485.

6. BIBLIOGRAFÍA

- [1] S. Bayarri, M. Fernández and M. Pérez. "Virtual Reality for Driving Simulation". Communications of ACM (Mayo 1996).
- [2] S. Bayarri, M. Fernández & Coma I. "Virtual Reality in Driving Simulation". Encyclopedie of Computer Science & Technology. Vol 38 pp 340-360. Ed. Marcel Dekker. 1998.
- [3] J. Casillas y otros. "Implementación del sistema visual de un Simulador de Helicóptero". CEIG '99 Jaen 1999.
- [4] M. Kleiner, B-I. Dalenbäck, P. Svensson. "Auralization - An overview". J. Audio Eng. Soc., Vol. 41, No. 11, November 1993
- [5] M. Vorländer. "Auralization in Acoustics". International Acoustical Conference - EAA Symposium. 2006
- [6] Vorländer. Auralization Fundamentals of Acoustics, Modelling, Simulation, Algorithms, and Acoustic Virtual Reality. [Editorial] 2008
- [7] CIPIC HRTF database:
http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm
- [8] J.M. Jot, V. Larcher, and O. Warusfel. "Digital signal processing issues in the context of binaural and transaural stereophony". Proc. 98th Audio Eng. Soc. Convention, 1995.
- [9] H. Møller. "Reproduction of artificial-head recordings through loudspeakers". J. of the Audio Engineering Society, 37(1/2):30–33, 1989.
- [10] H. Møller. "Fundamentals of binaural technology". Applied Acoustics, 36:171–218, 1992.
- [11] SM. Kim and W. Choi. "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach". J. of the Acoustical Soc. of America, 117, 3657, 2005.
- [12] D.H. Cooper and J.L. Bauck, "Prospects for Transaural Recording", J. Audio Eng. Soc., 37 (1/2), 3-19, 1989.
- [13] B. Gardner. "Transaural 3D audio". Technical Report 342, M.I.T. Media Lab Perceptual Computing, July 1995.
- [14] T. Lentz. "Binaural technology for virtual reality". Doctoral thesis, RWTH Aachen University, Germany. 2007.
- [15] V. Pulkki. "Virtual sound source positioning using vector base amplitude panning". J. of the Audio Eng. Soc., 45(6):456–466, June 1997.
- [16] V. Pulkki. "Spatial Sound Generation and Perception by Amplitude Panning Techniques". Doctoral thesis, Helsinki University of Technology, 2001.

- [17] M.A. Gerzon. "Ambisonics in multichannel broadcasting and video". J. of the Audio Engineering Society, 33(11):859–871, 1985.
- [18] D.G. Malham and A. Myatt. "3-D sound spatialization using ambisonic techniques". Computer Music Journal, 19(4):58–70, 1995.
- [19] J Daniel, R Nicol, S Moreau. "Further investigations of high order ambisonics and wavefield synthesis for holophonic Sound Imaging". J Audio Eng. Soc., 2003.
- [20] A.J. Berkhout, D. de Vries, and P. Vogel. "Acoustic control by wave field synthesis". J. of the Acoustical Soc. of America, 93(5):2764–2778, may 1993.
- [21] E.M. Wenzel, S.H. Foster, and F.L. Wightman, "Realtime Digital Synthesis of Localized Auditory Cues over Headphones", Proc. ICASSP, 1989.
- [22] OpenAL API 1.1: <https://connect.creativelabs.com/openal>
- [23] Microsoft Direct Sound 3D: [http://msdn.microsoft.com/en-us/library/bb219818\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/bb219818(VS.85).aspx)
- [24] Creative Technology Ltd. Environmental Audio Effects (EAX) API 5.0: <http://www.soundblaster.com/technology/welcome.asp?j1=eax>
- [25] Java 3D API: <http://java.sun.com/javase/technologies/desktop/java3d/>
- [26] Virtual Reality Modeling Language (VRML): <http://www.web3d.org/x3d/vrml/index.html>
- [27] Extensible 3D (X3D): <http://www.web3d.org/x3d/>
- [28] MPEG-4 AudioBIFS: ED Scheirer, R Vaananen, J Huopaniemi. "AudioBIFS: Describing audio scenes with the MPEG-4 multimedia standard". IEEE Trans. on Multimedia, 1999
- [29] J Segura; L Vera; A Barba; R Lakatis; A Giménez; M Fernández; R Cibrián; S Cerdá; J Romero; "Estudio de parámetros de calidad sonora de señal música y habla auralizada en una sala de uso múltiple". Cádiz, TecniAcústica 2009.