

SONIDO VIRTUAL O CÓMO MOVER UNA FUENTE MONOFÓNICA POR EL ESPACIO

Màrius Valles Segura, Néstor Andreu Ramírez, Enric Monte Moreno

marius_valles@menta.net, nestor_andreu@yahoo.es, enric@gps.tsc.upc.es

ABSTRACT

En este artículo se presenta un posible modelo capaz de sintetizar sonido virtual a partir de una fuente monofónica manteniendo la ilusión de espacialidad. Este modelo se basa en una descripción simplificada en el dominio temporal del comportamiento de las ondas sonoras (teniendo en cuenta el filtrado que el cuerpo realiza sobre ellas) y en las características más relevantes de la percepción humana concernientes a su localización (difracción, retardo y reflexión del sonido).

1. INTRODUCCIÓN

La reproducción del sonido en 3D está incrementando su importancia dentro del mundo científico, comercial y del ocio, por lo que la demanda de sistemas capaces de operar en tiempo real aumenta cada vez más.

El oído humano es capaz de situar en el espacio una fuente de sonido. Para lograrlo tiene en cuenta las reflexiones y el filtrado que la configuración anatómica de cada oyente realiza. La función de transferencia que relaciona el sonido que llega al tímpano con el emitido por la fuente se denomina HRTF (*Head Related Transfer Function*). Ésta puede ser sintetizada desarrollando un modelo estructural relativamente simple cuyos parámetros están estrictamente relacionados con las características fisiológicas de cada individuo.

La correcta simulación de la HRTF permite crear la ilusión de espacialidad en el cerebro, permitiendo que pueda localizar la procedencia de una onda sonora determinada.

2. PERCEPCIÓN ESPACIAL DEL SONIDO

Cuando un sonido proviene de algún lugar del espacio, una persona es capaz de localizar su procedencia en azimut, elevación y distancia. Sin duda, la estimación de su posición en azimut es mucho más precisa que el resto. La *Teoría Duplex* de Lord Rayleigh da cuenta de los principales factores que intervienen. Principalmente se habla de la diferencia de tiempo interaural o ITD (*interaural time difference*) y de la diferencia de nivel interaural o ILD (*interaural level difference*). Antes de definirlos, es

necesario establecer un sistema de coordenadas que permita especificar la posición de la fuente respecto al oyente: un sistema esférico centrado en la cabeza del oyente. La figura 2.1 muestra gráficamente su comportamiento:

2.1 Localización en el plano de azimut

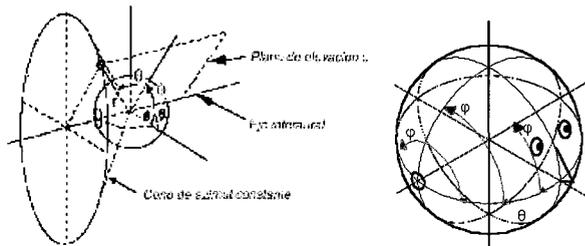
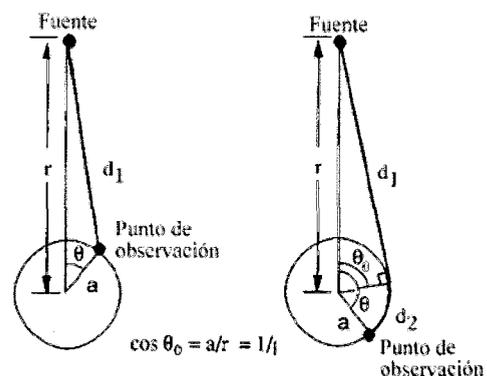


Figura 2.1. Sistema de coordenadas polar interaural. Se puede observar que la distinción entre delante y detrás se realiza mediante la elevación (j) y no azimut (q).

Por lo general, el sonido llega en momentos diferentes a cada oreja ya que debe recorrer caminos distintos, esta diferencia de tiempos es el ITD y se puede calcular de manera geométrica de la siguiente manera (Figura 2.2):



$$d_1 = \sqrt{r^2 + a^2 - 2ra \cos \theta} \quad d_1 + d_2 = \sqrt{r^2 - a^2} + a(\theta - \theta_0)$$

Figura 2.2. El sonido puede llegar de forma directa al punto de observación (d_1) o de manera indirecta ($d_1 + d_2$).

Así pues, esta diferencia de tiempo $\Delta \tau$ viene dada por:



$$\Delta\tau = \frac{c\Delta t}{2\pi a} = \begin{cases} \frac{1}{2\pi}(\sqrt{\rho^2 - 2\rho \cos(\theta) + 1} - \rho) & \text{para } 0 \leq \theta \leq \theta_0 \\ \frac{1}{2\pi}(\theta - \theta_0 + \sqrt{\rho^2 - 1} - \rho) & \text{para } \theta_0 \leq \theta \leq \pi \end{cases} \quad (1)$$

donde,

$$\rho = r/a \quad (2)$$

$$\theta_0 = \sin^{-1}(1/\rho) \quad \text{para } \rho \geq 1$$

Existe, también, una aproximación del ITD muy realista calculada de manera empírica. Su valor oscila entre dos valores diferentes, dependiendo si es a bajas frecuencias o a altas, de la siguiente manera (siendo k el número de onda ()):

$$ITD_{baja} = \frac{3a}{c} \cdot \text{sen}\theta, \quad \text{para } ka \ll 1 \quad (3)$$

$$ITD_{alta} = \frac{2a}{c} \cdot \text{sen}\theta, \quad \text{para } ka \gg 1 \quad (4)$$

Estos dos valores de ITD son sólo valores asintóticos, en realidad el ITD tiene una pendiente decreciente en función de la frecuencia. La figura 2.3 muestra cómo varía el ITD real según la frecuencia para una fuente situada a $\theta = 40^\circ$ y dichos valores asintóticos.

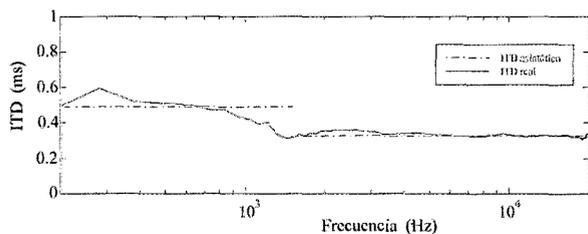


Figura 2.3. Variación del ITD según la frecuencia con cabeza esférica.

La cabeza produce una zona de sombreado acústico en el lado opuesto al que le llega el sonido directo, es decir, existe una diferencia en el nivel sonoro de las señales en ambos oídos (ILD). Lord Rayleigh solucionó la ecuación de onda en una esfera, de esta manera se puede determinar exactamente la medida de esta difracción producida para un frente incidente de ondas planas (fuente situada en el infinito). Como es de esperar, el ILD es altamente dependiente de la frecuencia. En la figura 2.4 se muestra la respuesta de una esfera rígida de radio a según la solución de Rayleigh. En ella se observa la relación existente entre la intensidad de la onda en el punto de incidencia perpendicular y la intensidad en un punto situado a un ángulo β de éste.

Obviamente, el considerar la cabeza como una esfera es una aproximación que permite calcular rápidamente el ILD. En realidad, en la cabeza se producen numerosas reflexiones y difracciones que, en altas frecuencias, provocan un ILD aún mayor. Además, el ILD no sólo depende del ángulo y de la frecuencia sino también de la distancia a la que se encuentra

la fuente, sobretodo para distancias inferiores a 1m., Donde el ILD tiende a ser mayor y se comporta de un modo ligeramente distinto.

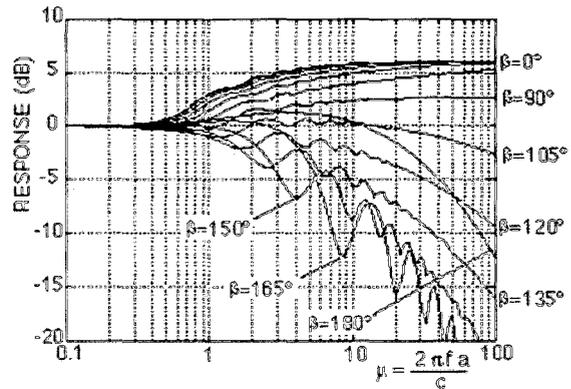


Figura 2.4. Respuesta de una esfera rígida de radio a para el cálculo del ILD.

Según la teoría Duplex, el ILD y el ITD son fenómenos complementarios. A frecuencias bajas ($\approx 1,5$ KHz) la diferencia de tiempos aporta información importante para la localización mientras que la diferencia de nivel es casi despreciable. Por otro lado, a partir de 3 KHz, es más importante el ILD que el ITD. Es por este motivo que el sistema de localización tiende a ser pobre cuando se intenta localizar un sonido que sólo tenga componentes frecuenciales en el margen de 1,5 KHz a 3 KHz.

2.2 Localización en el plano de elevación

Si se mantiene la aproximación de una cabeza esférica ideal, el ITD y el ILD no varían en función de la elevación. Sin embargo, el oído externo sí produce un filtrado que depende de la elevación, por lo que resulta de especial importancia para detectar la elevación de la fuente. Su configuración anatómica lo hace funcionar como una antena receptora, con cavidades resonantes que amplifican unas determinadas frecuencias y configuraciones geométricas que atenúan otras. Su respuesta depende del ángulo de incidencia del sonido, tal y como se muestra en la figura 2.5.

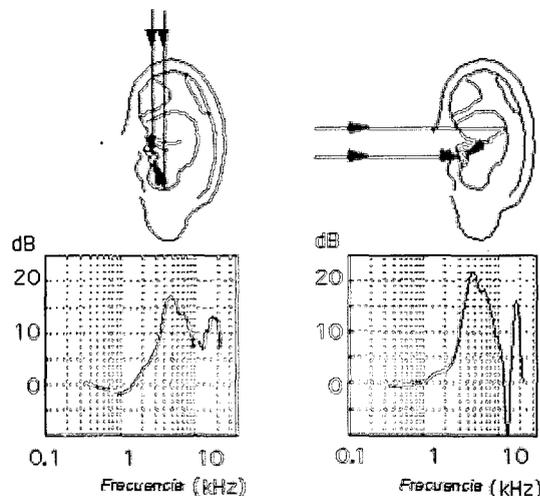


Figura 2.5. Respuesta frecuencial del oído externo con diferentes ángulos de incidencia.

Como se puede observar en la figura anterior, el sonido entra en el conducto auditivo de dos maneras diferentes: por el camino directo sin retardo y a través de reflexiones en el oído externo retardadas. A frecuencias bajas, básicamente, la energía de los diferentes caminos recorridos se suma en fase. Sin embargo, a altas frecuencias las señales que llegan retardadas no tienen por qué llegar en fase con la señal del camino directo y, por lo tanto, se pueden producir interferencias destructivas. La interferencia destructiva máxima tiene lugar cuando la distancia que recorre el sonido reflejado es: $d = \lambda/2$, siendo λ la longitud de onda, y generalmente se produce entre 6KHz y 16KHz, dependiendo de la morfología del oído del oyente. Además, cabe destacar que este pico mínimo es mayor cuando el sonido incide frontalmente que cuando lo hace desde arriba y, como la distancia d cambia en función del ángulo de incidencia, la posición de éste varía en función de la elevación. Y es precisamente esta variación de la frecuencia en función de la dirección la que el cerebro es capaz de reconocer para localizar la fuente sonora. Además, para una fuente situada en el plano de $\theta=0^\circ$ (plano vertical que divide la cabeza en dos mitades iguales), tanto el ILD como el ITD son nulos, de forma que ésta es la única forma de localizar la fuente.

2.3 Localización en distancia

La localización en distancia no depende de unos parámetros tan fácilmente caracterizables como el ITD o el ILD. Es necesario el uso de diversas técnicas como, por ejemplo, añadir algún eco y cola reverberante al sonido tratado (de forma que la impresión es mucho más natural, puesto que cualquier sonido que se escucha en la realidad presenta estos fenómenos) o modificar el volumen al que la fuente suena.

Hay que tener en cuenta que, cuando una fuente suena muy cercana a un oído, proporcionalmente existe una distancia bastante mayor que la separa del otro. Esto provoca que exista un ILD mucho mayor del que causaría la cabeza por sí sola. Este incremento de ILD es más independiente de la frecuencia, pues está causado principalmente por la atenuación en distancia, y es determinante para localizar fuentes cercanas al oyente, especialmente para rangos inferiores a un metro.

Otro fenómeno útil es el propio movimiento de la cabeza para ayuda a determinar la distancia. Si la fuente está en un punto cercano, la posición en azimut en la que se localiza variará según se mueva el oyente. En cambio, si la fuente se halla lejos, su azimut apenas variará. Es probable que el cerebro use este fenómeno, conocido como paralaje, para precisar mejor la distancia a la que se encuentra la fuente.

3. SÍNTESIS DE SONIDO VIRTUAL

El modelo que se propone es un modelo simple de procesado de señal capaz de sintetizar sonido binaural a

partir de una fuente monoaural. Éste mantiene una correspondencia unívoca con los hombros, la cabeza y el oído externo del oyente, cada elemento aporta una característica temporal única a la respuesta impulsional global. El principal motivo por el que se presenta el modelo en el dominio temporal es porque el sonido llega al oído interno a través de múltiples caminos. Señales que llegan por caminos de longitud diferente interactúan de una manera mucho más clara en el dominio temporal que en el frecuencial. El esquema del modelo general se presenta en la figura 3.1. Principalmente, el modelo de la cabeza y los hombros proporcionan información para localizar en azimut y en rango, generando el ITD e ILD necesarios para cada posición que se procese, mientras que el modelo del oído externo permite la localización en elevación. A continuación se detallará individualmente cada uno de sus componentes.

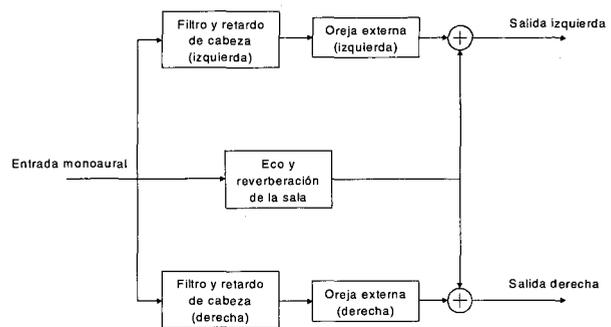


Figura 3.1. Componentes del modelo

3.1 Modelado de la cabeza

Es posible sintetizar el ILD con una función simple de un polo y un cero, consiguiendo así modelar el efecto de sombreado acústico que causa la cabeza. La siguiente ecuación muestra dicho filtro de sombreado acústico:

$$H_{sa}(\omega, \beta) = \frac{1 + j \frac{\alpha(\beta)\omega}{2\omega_0}}{1 + j \frac{\omega}{2\omega_0}} \quad \text{para } 0 \leq \alpha(\beta) \leq 2 \quad (5)$$

con valor de la constante $\omega_0 = c/a$.

La frecuencia normalizada referida a la constante ω_0 es $\mu_0 = \omega_0 a/c = 1$. El coeficiente α depende del ángulo de incidencia β , controlando así la posición donde se encuentra el cero. Si $\alpha=2$ existe un pico a altas frecuencias de 6dB, mientras que para $\alpha<1$ hay una zona de corte. Para coincidir con el patrón de las curvas mostradas en la figura 2.4, α debe depender de β de la siguiente manera:

$$\alpha(\beta) = \left(1 + \frac{\alpha_{\min}}{2}\right) + \left(1 - \frac{\alpha_{\min}}{2}\right) \cos\left(\frac{\beta}{\beta_{\min}} 180^\circ\right) \quad (6)$$

donde las constantes que producen una buena aproximación son:

$$\alpha_{\min} = 0.1$$

$$\beta_{\min} = 150^\circ$$

Por otro lado, para sintetizar el ITD basta con modelar un elemento retardador correspondiente con la ecuación 2. De esta manera, si se encadenan ambos filtros se obtiene la aproximación $\hat{H}(\infty, \mu, \beta)$.

Además, nótese que H_{sa} introduce el siguiente retardo de grupo:

$$\tau_g = \frac{1 - \alpha(\beta)}{2\omega_o} = \frac{a}{2c}(1 - \alpha(\beta)) \quad (7)$$

el cual, aumenta el tiempo $\Delta\tau$ para frecuencias bajas ($\mu < 1$), de forma que lo acerca más a los resultados observados por Khun [3]. De hecho, para $\beta=0^\circ$ este retardo de grupo añade exactamente un retardo adicional del 50%. La figura 3.2 muestra la respuesta frecuencial del filtro H_{sa} y la respuesta temporal del modelo retardador. Se puede observar como las curvas de respuesta frecuencial se aproximan de manera satisfactoria a las teóricas mostradas en la figura 3.2:

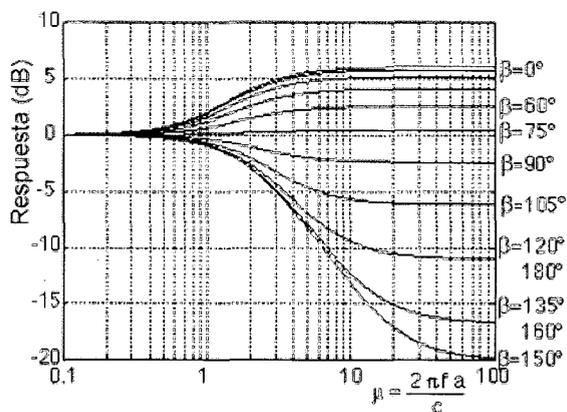


Figura 3.2. Respuesta frecuencial del filtro de un polo y un cero que aproxima la respuesta de una esfera rígida.

3.2 Modelado del oído externo

El oído externo se simula mediante un modelo de propagación multicamino que simula los eventos observados en diversos experimentos, cuyo esquema se puede observar en la figura 3.3:

Realizando tests auditivos se ha podido determinar que cinco eventos son suficientes para este modelo y que su amplitud se puede aproximar como constante sea cual sea la dirección de la que la que provenga el sonido. Lo único

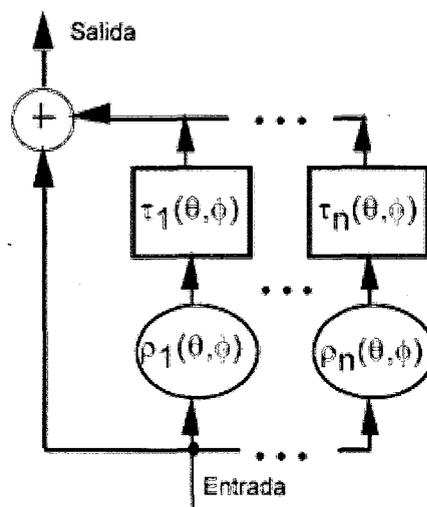


Figura 3.3. Esquema del modelo del oído externo. Donde ρ_k y τ_k son, respectivamente la amplitud y el retardo de los eventos.

que varía son sus retardos, que se pueden aproximar por la siguiente ecuación:

$$t_k(q, f) = A_k \cos(q/2) \sin(D_k(90^\circ - f \cos q)) + B_k \quad (8)$$

Se da una cierta variación entre distintos individuos en la morfología de la oreja externa. Para adaptarse a ellas solamente hace falta variar los parámetros D_k . A continuación se presenta una tabla con los valores de los parámetros y las amplitudes para los cinco eventos y para un individuo concreto. A_k y B_k se dan en muestras para una frecuencia de muestreo de 44100 Hz:

k	ρ_k	A_k	B_k	D_k
1	0.5	1	2	1
2	-1	5	4	0.5
3	0.5	5	7	0.5
4	-0.25	5	11	0.5
5	0.25	5	13	0.5

3.3 Eco y reverberación de la sala

Para poder externalizar bien el sonido y así, localizar en rango, es necesario crear la impresión de encontrarse en una sala. Para ello, se deben generar ecos y reverberaciones del sonido original y añadirlos al sonido ya tratado. Se ha encontrado que un modelo rápido, pero suficientemente preciso, consiste en crear un filtro que genere un solo eco y que, además, disponga de unos cuantos lazos de realimentación para generar la reverberación. El número de coeficientes, así como su valor y retardo, dependerán del tipo de sala o efecto que se quiera conseguir (de forma aproximada), y la mejor manera de encontrarlos es de forma empírica. Cuantos más lazos de realimentación se usen, más rica resulta la reverberación, aunque con cinco o seis ya es posible crear reverberaciones convincentes.

Para hacer la reverberación más realista, hay que hacer un filtrado paso-bajo a la señal antes de que pase por los lazos, de forma que se simula el hecho de que los materiales absorben mejor las altas frecuencias que las bajas.

4. IMPLEMENTACIÓN DEL MODELO

Hasta este punto, se ha presentado el modelo de manera teórica. Sin embargo, si se desea implementarlo sobre algún lenguaje de programación se deben tener diversas consideraciones en cuenta. La principal diferencia radica en la posibilidad de realizar las simulaciones con fuentes dinámicas, es decir, de definir una trayectoria sobre la cual se debe mover la fuente. La figura 4.1 muestra el esquema para una implementación práctica del modelo:

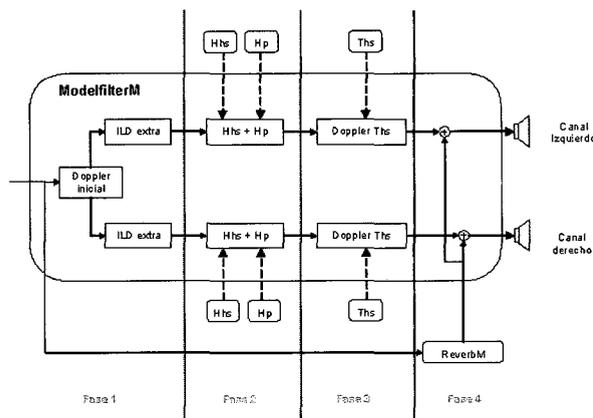


Figura 4.1. Modelo para la implementación práctica del sistema.

La señal de entrada se divide en bloques para acelerar el procesado, puesto que hay muchos cálculos que sólo es necesario realizarlos una vez por bloque (por ejemplo, los coeficientes de los filtros o el cálculo de los retardos). Cuanto mayores son los bloques, mayor es la rapidez de procesado, pero también es mayor la latencia del sistema y menor su precisión espacial. El compromiso, sin embargo, es fácil de encontrar, puesto que la eficiencia prácticamente no mejora para bloques mayores de 300 muestras, y con este tamaño tanto la latencia como la resolución espacial son totalmente aceptables.

Como se ve, el procesado se realiza en cuatro fases encadenadas, que a continuación se explican más detalladamente:

Fase 1: Aquí se calcula el efecto Doppler relacionado con el movimiento de la fuente respecto del centro de la cabeza del oyente. También se introduce el ILD extra que aparece cuando la fuente se encuentra cerca del oyente. El efecto Doppler se calcula mediante repetición o supresión de muestras, según convenga, para obtener mayor velocidad. Otros

sistemas más precisos que evitan efectos de aliasing se han descartado por el enorme incremento de tiempo de cálculo que representan.

Fase 2: En esta fase se realiza el filtrado correspondiente tanto a la sombra acústica como a las orejas exteriores. El primero (Hhs) es de implementación inmediata. Sin embargo, la realización del segundo (Hp) tiene la dificultad de que los eventos de la respuesta impulsional no caen normalmente justo en el tiempo de inicio de una muestra. Para subsanar esto, se divide un evento en dos muestras de la respuesta impulsional de forma proporcional, tal y como indica la figura 4.2

Posición teórica de la delta

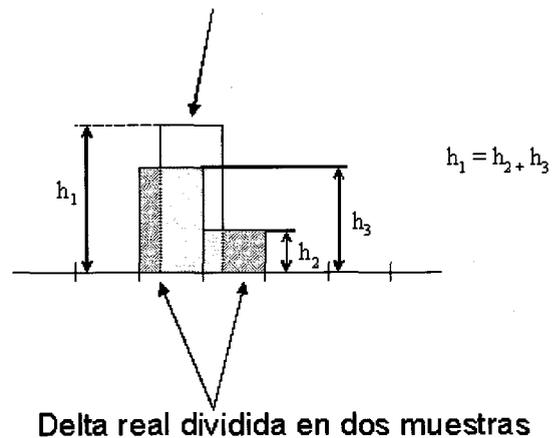


Figura 4.2. Separación de un evento (delta) en dos muestras cuando su inicio no coincide exactamente con el de ninguna muestra.

Fase 3: Aquí se generan los retardos entre canales derecho e izquierdo de forma que se adapten al movimiento de la fuente. Para ello es necesario volver a aplicar efectos Doppler que, aunque sus efectos en frecuencia son prácticamente inaudibles, comprimen o expanden las señales para que se adapten a dichos retardos.

Fase 4: En esta última fase se añade la salida del módulo de reverberación, obteniendo la salida binaural.

Este sistema está pensado para que su salida se conecte directamente a unos auriculares, puesto que las señales calculadas para los canales derecho e izquierdo son exactamente las que deben recibir las respectivas orejas. Es posible, sin embargo, obtener el mismo resultado con dos altavoces, aunque el sistema se complica porque el altavoz de un lado tiene que emitir a la vez su señal y la que debe cancelar la proveniente del otro altavoz. Este sistema tendría el problema añadido de que dicha cancelación sólo se obtendría en una zona muy pequeña de la sala (figura 4.3):

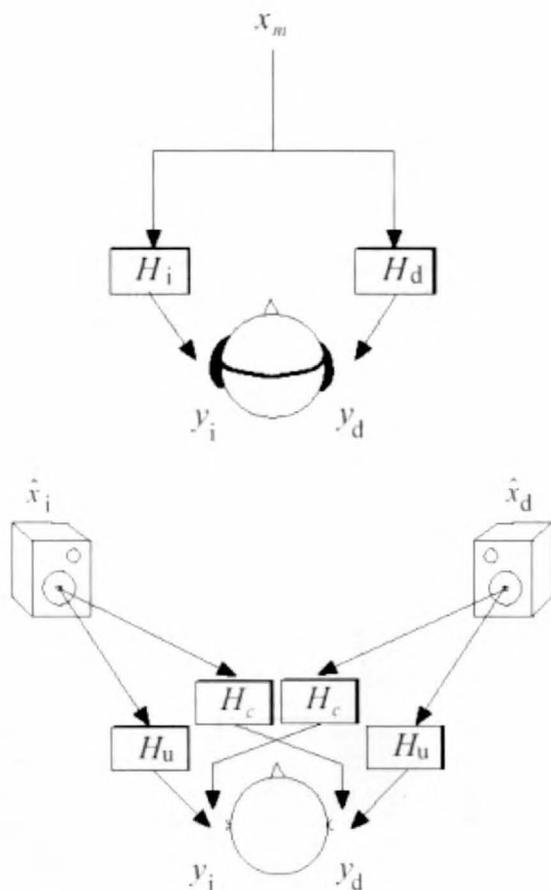


Figura 4.3. Métodos de reproducción: con auriculares (filtrado independiente de la señal) o con dos altavoces (filtrado útil de la señal y filtrado cancelador de crosstalk H_c).

5. CONCLUSIONES

Se ha presentado un modelo simple de procesado de señal para la síntesis de sonido virtual. Dicho modelo, contiene componentes que procesan, de manera independiente, los elementos que hacen posible la localización de una fuente en el espacio: el ILD y ITD para la localización en azimut, el modelado del oído externo para la elevación y la reverberación para el rango. Los diferentes experimentos acústicos realizados han demostrado que con este modelo se puede simular de manera convincente la percepción espacial del sonido. Además, la simplicidad de los filtros usados y el planteamiento en sí del modelo, lo convierten en un sistema ideal para una implementación a tiempo real.

Otra ventaja añadida de este sistema es que se puede adaptar a cada usuario, puesto que permite personalizar los parámetros más relevantes para la percepción (los referidos, fundamentalmente, a la configuración anatómica del oído externo). De este modo, se incrementa la sensación de realismo.

En definitiva, la sencillez, la eficiencia y su versatilidad hacen que el sistema presentado en este artículo resulte óptimo para usos tanto científicos como comerciales, ya que permite una simulación rápida y lo suficientemente precisa de sonido 3D.

6. REFERENCIAS

- [1] Brown, C. Phillip & Duda, Richard O.; «A Structural Model For Binaural Sound Synthesis», 1998.
- [2] Brown, C. Phillip & Duda, Richard O.; «An Efficient HRTF Model For 3-D Sound», 1997.
- [3] Kuhn, G. F.; «Model for the interaural time differences in the azimuthal plane,» *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157-167, July 1977.
- [4] Hupaniemi, Jyri; «Virtual Acoustic and 3D sound in multimedia signal processing», Helsinki University of Technology (Laboratory of Acoustics and Audio Signal Processing), Report 53, 1999
- [5] Algazi, V. Ralph & Duda, Richard O.; «Approximating the head-related transfer function using simple geometric models of the head and torso», *J. Acoust. Soc. Am.*, vol. 112 no.1, Nov 1992.
- [6] Algazi, V. Ralph et al.; «Structural Composition and decomposition of HRTFs», IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.
- [7] Algazi, V. Ralph et al.; «The CIPIC HRTF database», IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001
- [8] Hartmann, William M.; «How we localize sound», *Physics Today*, Npv 1999.

AUTORES



Marius Valles Segura. Desde 1998 ha estado cursando Ingeniería Superior de Telecomunicaciones, y actualmente está realizando el Proyecto de Fin de Carrera. Ha participado en varias ediciones del Concurso Nacional de robots de la UPC. Ha realizado un curso de acústica y ya se encuentra asociado al colegio de ingenieros de telecomunicaciones.



Néstor Andreu Ramírez. Desde 1998 ha estado cursando Ingeniería Superior de Telecomunicaciones, y actualmente está realizando el Proyecto de Fin de Carrera. Ha participado en varias ediciones del Concurso Nacional de robots de la UPC. Ha realizado un curso de acústica y ya se encuentra asociado al colegio de ingenieros de telecomunicaciones.



Enric Monte Moreno. Se graduó y se doctoró en Ingeniería de Telecomunicación por la Universidad Politécnica de Catalunya (UPC) en 1987 y 1992 respectivamente. Vinculado al departamento de Teoría de Señal y Comunicaciones desde el año 1989, actualmente ejerce el cargo de profesor titular. Su interés actual se centra en el tratamiento de señal de voz.