

Modelo basado en evidencias perceptuales para simular respuestas impulsivas de recintos

Fabián C. Tommasini¹, Oscar A. Ramos²

Centro de Investigación y Transferencia en Acústica, CINTRA, UA del CONICET,
Universidad Tecnológica Nacional, Facultad Regional Córdoba,
Maestro M. López esq. Cruz Roja Argentina, X5016ZAA, Córdoba, Argentina
ftommasini@gmail.com, oramos@scdt.frc.utn.edu.ar

Resumen. Un recinto es caracterizado por su respuesta impulsiva. Se compone de: el sonido directo, las reflexiones tempranas (especulares y difusas) y las reflexiones tardías. Es esencial, para un sistema de realidad acústica virtual, simular de manera convincente dicha respuesta. El objetivo del presente trabajo es desarrollar un modelo que permita generar de manera eficiente la respuesta impulsiva de un recinto. Para alcanzar este propósito, se realizan simplificaciones basándose en evidencias perceptuales. Se utiliza el método de la fuente-imagen para obtener las reflexiones tempranas especulares; mejoras de filtros pasa-todo para las reflexiones tempranas difusas; mientras que, para las reflexiones tardías, se emplea una red de retardo realimentada.

Palabras clave. Respuesta Impulsiva de Recinto, Realidad Acústica Virtual, Auralización.

1 Introducción

La simulación acústica virtual, llamada también Espacio Acústico Virtual (*Virtual Acoustic Space*, VAS), permite sintetizar sonidos que al ser escuchados por un oyente, producirán en él, la sensación de presencia en el espacio simulado. Hablamos de Realidad Virtual (*Virtual Reality*, VR) si este proceso se realiza en tiempo real de manera que el usuario pueda interactuar con el entorno modelado. Al proceso de generar las señales acústicamente correctas para dicho fin, se lo conoce como *auralización*.

El término auralización fue introducido en 1993 por Kleiner et al. [1] donde indica que: “la vivencia auditiva que un oyente experimentaría en un espacio dado (real o imaginario) podrá ser evocada si, mediante modelos físico-matemáticos, se sintetizan y reproducen en los tímpanos del oyente las ondas sonoras que se originarían en el espacio modelado”.

En un sistema de realidad acústica virtual, tres subsistemas deben ser modelados:

¹ Becario doctoral del CONICET

² Miembro del CONICET

la fuente sonora, el recinto y el medio de propagación y el oyente.

En este trabajo se presenta el modelo del recinto. Para cada configuración fuente-receptor, existe una función de transferencia o respuesta impulsiva que lo caracteriza. Estrictamente hablando, modelar el campo sonoro en un recinto consiste en resolver las ecuaciones diferenciales de la onda en el espacio y el tiempo. Pero la potencia de cálculo requerida es enorme y su aplicación se restringe a pequeños recintos rectangulares y para bajas frecuencias [1].

Nuestro abordaje es diferente. Basándonos en evidencias perceptuales se consigue simplificar los modelos físico-acústicos, obteniéndose así modelos que son “perceptualmente aceptables” para ser ejecutados en tiempo real en una computadora personal.

Asumiendo que existe un camino sin obstrucción entre la fuente y el oyente, éste escuchará primero el *sonido directo*, seguido por las *reflexiones tempranas* (especulares y difusas) sobre las superficies más cercanas. Luego de algunos milisegundos (varía de 60 a 100 ms en la mayoría de los recintos), la densidad de las reflexiones aumenta, y su comportamiento se asemeja a un proceso aleatorio e independiente de la ubicación del receptor dentro del recinto. Esto se llama *reflexiones tardías* o *cola de reverberación*.

Un parámetro importante y que caracteriza un recinto es su tiempo de reverberación. Se define como el tiempo que tarda un sonido, después de que es interrumpido, en atenuarse 60 dB en relación al nivel inicial. El tiempo de reverberación varía con la frecuencia y, en general, es mayor para las bajas frecuencias (hasta 500 Hz), disminuyendo considerablemente para frecuencias mayores.

Las direcciones de las reflexiones tempranas son cruciales para que una persona pueda localizar con precisión una fuente sonora; mientras que las reflexiones tardías se asemejan a un proceso estocástico cuyas direcciones son irrelevantes pero claves para la percepción de las dimensiones, la calidad acústica del recinto y de la distancia de la fuente sonora [2].

Basándonos en estas evidencias, utilizamos para modelar las reflexiones tempranas el método de la fuente-imagen (*Image-Source Method*, ISM), que permite establecer con precisión la dirección de los rayos sonoros que arriban al receptor. La desventaja del método es que asume que las reflexiones son totalmente especulares, dejando de lado las reflexiones difusas. Para salvar esta deficiencia, se usan filtros pasa-todo de Schroeder mejorados. Finalmente, para las reflexiones tardías o cola de reverberación se utiliza una red de retardo realimentada (*Feedback Delay Network*, FDN).

2 Implementación

En esta etapa y con la finalidad de probar, depurar y optimizar los algoritmos se ha utilizado Matlab y se realizaron algunas simulaciones en Simulink. En una segunda etapa, se migrarán los algoritmos a código C++, más indicado para la ejecución en tiempo real. De todas maneras, ya se ha avanzado en el diseño de objetos y métodos eficientes.

En la Fig. 1, se muestra un esquema de las diferentes partes del modelo propuesto.

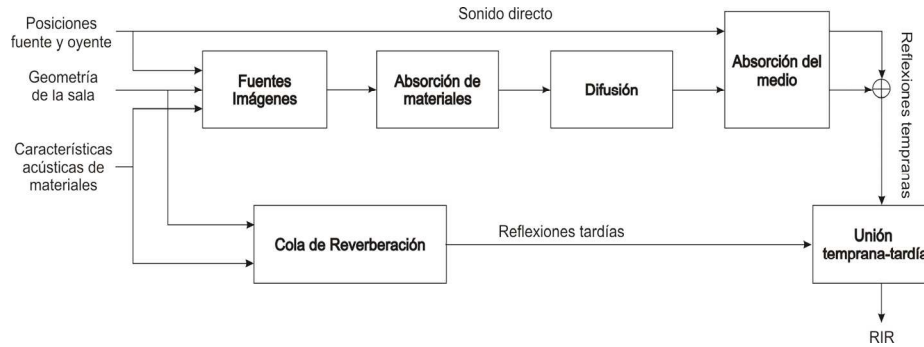


Fig. 1. Esquema del modelo para obtener una RIR simulada

Las salidas del bloque de las fuentes-imágenes, alimentan los filtros de absorción de los materiales y luego, los filtros de difusión y el de absorción del aire. El sonido directo es procesado solo por el filtro de absorción del aire. La RIR se consigue empalmando las reflexiones tempranas y las tardías, generadas en el bloque “cola de reverberación”. La implementación de los filtros de materiales y el de absorción del aire puede ser consultada en [3].

2.1 Reflexiones Tempranas Especulares

El método de la fuente-imagen es determinístico y está basado en principios de acústica geométrica que considera a una onda sonora como un rayo. El rayo representa la trayectoria de la onda desde una fuente sonora hasta el receptor.

El campo sonoro total en la posición del oyente R está conformado por el rayo directo, originado por la fuente real S (el que recorre el mínimo camino entre la fuente y el oyente), y por un conjunto de rayos originados en las fuentes imágenes o virtuales (*virtual source*, VS) (Fig. 2).

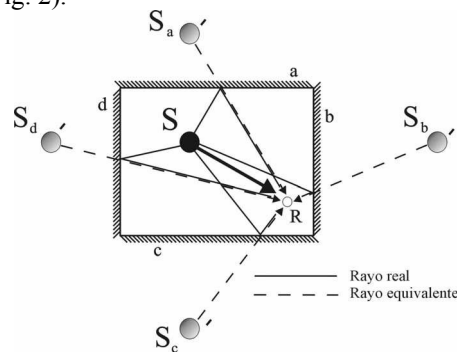


Fig. 2. Método de la fuente-imagen. Los rayos reales se reemplazan por rayos que emiten las fuentes imágenes.

Este método asume que: las superficies (paredes, piso, techo, entre otros) son polígonos que pertenecen a planos infinitos y que tienen una cara reflejante (la del interior del recinto) y una cara opaca no reflejante; que la fuente es omnidireccional y puntual, al igual que el receptor.

Si el recinto tiene n superficies, entonces hay n posibles VS de primer orden (S_1, S_2, \dots, S_n). Cada una de estas (VS padres), a su vez, dan origen a $(n - 1)$ VS de orden superior (VS hijas). Para un orden dado i , el número de VS crece exponencialmente de la forma $n(n - 1)^{i-1}$ [4]. Como se ha podido advertir, el orden de una VS está dado por la cantidad de veces que el rayo sonoro ha sido reflejado en la superficie antes de alcanzar al receptor.

La ubicación de un plano que contiene a una superficie reflejante está determinada por dos parámetros: la normal unitaria \hat{n} al plano y su distancia al origen de coordenadas p . La posición de una VS S_v se encuentra haciendo:

$$S_v = S_r + 2d\hat{n} \quad (1)$$

con

$$d = p - S_r\hat{n} \quad (2)$$

donde d es la distancia de la VS al plano; S_r y S_v son las coordenadas de la VS padre y la VS hija respectivamente (Fig. 3a).

El algoritmo es recursivo debido a que cada una de las VS originadas en una superficie, genera nuevas VS a partir de las superficies restantes. De esta manera el conjunto de VS forman una estructura de datos tipo árbol n -ario. La organización de los datos en este tipo de estructura facilita el procesamiento que se ve a continuación.

No todas las VS contribuyen al campo sonoro total. Cada VS debe cumplir tres requisitos: a) validez, b) proximidad, y c) visibilidad. Por razones de espacio, no se explica aquí los dos primeros requisitos. En un artículo previo [5] están desarrollados con suficiente extensión los puntos a) y b). Se describe a continuación el requisito de visibilidad para VS de orden superior.

En [6] se discute un método basado en determinar receptores virtuales. Sin embargo, el costo computacional es importante. La solución que proponemos consiste en reconstruir la trayectoria del rayo desde el receptor a la fuente y verificar que la intersección del mismo con las superficies pertenecen a las superficies que originaron las VS (Fig. 3b).

Partiendo del árbol con las VS, se calcula el punto de intersección de una VS de orden superior (S_2) con el plano de la superficie que la generó (línea de puntos gruesos). Si el punto (p_1) está dentro de los límites de la superficie, se sigue verificando con la VS padre (S_1) y se comprueba que el punto (p_2) de intersección con el plano que la originó, pertenezca a la superficie respectiva. De esta manera se continúa hasta llegar a las VS de primer orden (que son verificadas en el criterio de visibilidad de primer orden). Si todos los puntos de intersección están dentro de los límites de las superficies respectivas, entonces la VS de orden superior cumple con el segundo criterio de visibilidad.

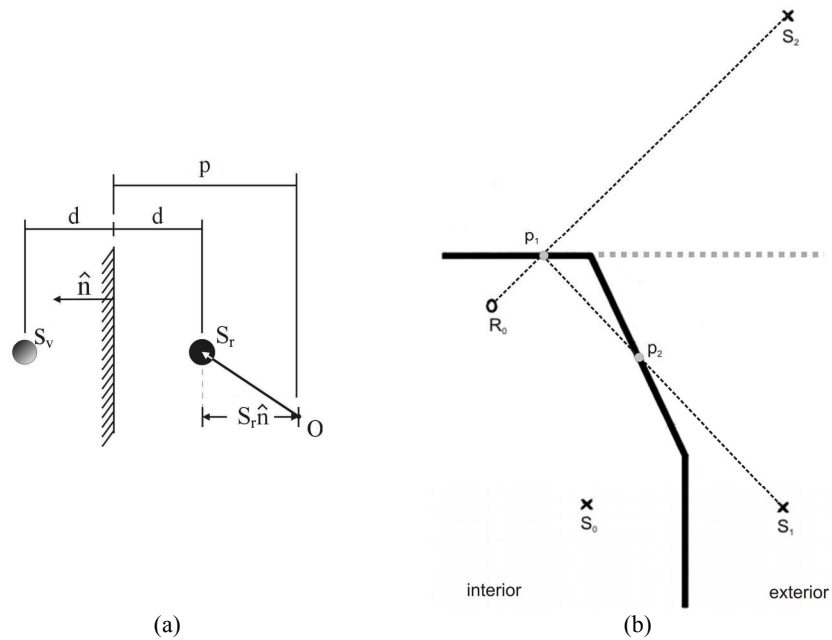


Fig. 3. Cálculo de ubicación de una VS (a) y requisito de visibilidad de orden superior (b).

Hay que tener en cuenta que solo las VS que son visibles contribuyen al sonido que llega al receptor. Sin embargo, las VS que no son visibles no son descartadas ya que alguna de sus VS hijas pueden serlo. Una VS es descartada solo cuando no cumple los requisitos de validez o proximidad. En la Fig. 4 se muestran las reflexiones tempranas especulares generadas por el modelo más el sonido directo. Las amplitudes disminuyen proporcionalmente con la inversa de la distancia de las VS al receptor.

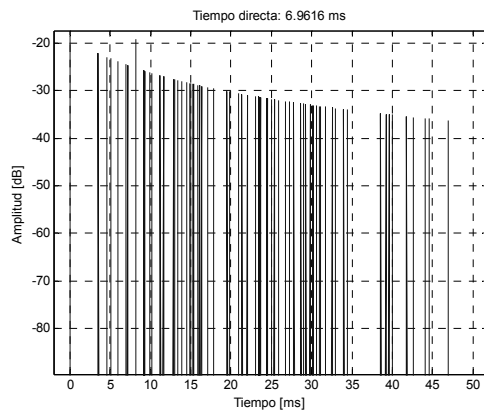


Fig. 4. Reflexiones tempranas especulares de tercer orden

2.3 Reflexiones Tempranas Difusas

En la realidad, al reflejarse un rayo en una superficie, simultáneamente con la reflexión especular se producen reflexiones difusas. Éstas tienen la característica principal de propagarse en todas las direcciones de manera estocástica.

Para modelar este comportamiento se utiliza una estructura en cascada de 8 filtros pasa-todo de Schroeder [7] mejorados. La mejora está basada en una observación de Pellegrini [8] que advirtió que al utilizar estos filtros se producía una “coloración” no deseada. En su opinión, la utilización de filtros pasa-todo complementarios, pueden atenuar este efecto.

Se propone la utilización de filtros con la siguiente función de transferencia:

$$H(z) = \frac{-\alpha + z^{-k} + \alpha\beta z^{-m} - \beta z^{-(m+k)}}{1 - \alpha z^{-k}} \tag{3}$$

donde $\alpha = e^{k\gamma}$ ($0 < \alpha < 1$), k es el número de polos, m es el retardo en muestras del filtro complementario (m debe ser múltiplo de k), $\gamma = \frac{\ln(\sqrt{1/2})}{k+1}$, $\beta = \frac{1-\alpha^2}{\alpha} e^{\gamma m}$.

La respuesta impulsiva de la ecuación (3) se muestra en la Fig. 5a en donde se puede ver que, después de un tiempo fijado por m , la respuesta abruptamente se aproxima a cero.

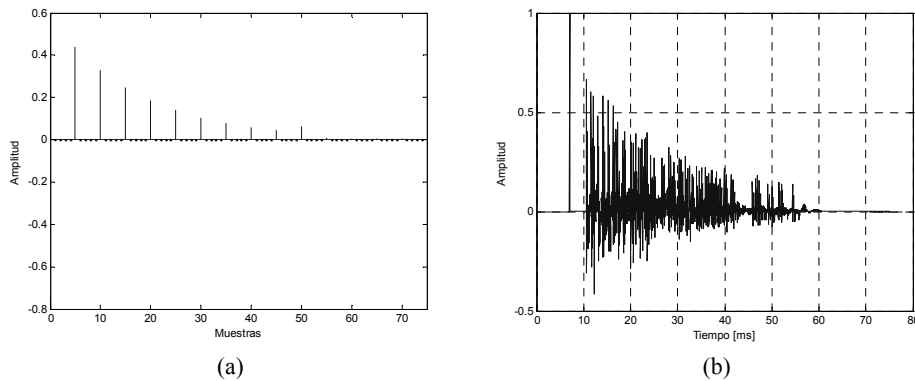


Fig. 5. Respuesta impulsiva del filtro de difusión (a); y en (b) las reflexiones tempranas especulares y difusas.

El valor de k determina, por un lado, la longitud de la respuesta del filtro y por el otro, de manera inversa, la densidad de los picos. Para conciliar estas dos variables, se utilizó k para determinar la longitud y se realizó una interpolación lineal a la respuesta impulsiva para lograr una mayor densidad. Las reflexiones tempranas especulares (Fig. 4) alimentan el banco de filtros propuestos, lográndose lo mostrado en Fig. 5b. Se utilizó k con un valor de 5 y m como 10 veces k .

2.4 Reflexiones Tardías

Las reflexiones tardías o cola de reverberación, se asemejan a un ruido aleatorio que se atenúa exponencialmente con el tiempo y determinan el tiempo de reverberación. Además, como ocurre en recintos reales, el tiempo de atenuación debe variar con la frecuencia. Debido a estas evidencias, no es necesario modelar las reflexiones tardías de manera precisa como lo hicimos con las tempranas. Es apropiado, entonces, la utilización de estructuras de filtros realimentados como es una red de retardo realimentada (FDN).

En 1991, Jot y Chaigne [9, 10] introdujeron la red de la Fig. 6 (llamada FDN de Jot), que permite especificar:

- la densidad de las reflexiones en el tiempo y en la frecuencia, con ausencia de coloración tonal; y
- el tiempo de reverberación y la respuesta en frecuencia deseada.

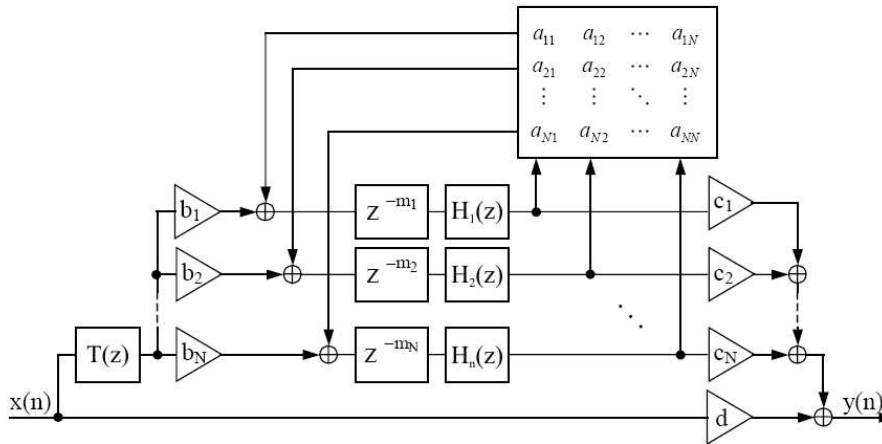


Fig. 6. Red de retardo realimentada de Jot.

Esta red, basada en la de Stautner y Puckette [11], tiene n líneas de retardo y una matriz de realimentación A . Cada coeficiente a_{ij} determina la proporción de la señal generada por la línea de retardo j que realimenta la línea de retardo i . La red de retardo es estable si A es el producto de una matriz unitaria y un coeficiente de ganancia g , donde $|g| < 1$.

Se construyó una red compuesta de 8 líneas de retardo y se utilizaron números coprimos para los valores de m (601; 691; 773; 839; 919; 997; 1061; 1129) [12]. La matriz A es de la familia de matrices de Householder, de la forma:

$$A = \frac{2}{n}uu^T - I \tag{4}$$

donde n es la cantidad de líneas de retardo, u es un vector de dimensión $n \times 1$ e I es la matriz identidad de orden $n \times n$.

Para conseguir que el tiempo de reverberación varíe con la frecuencia, se intercaló en cada línea un filtro que realiza esta función. En [10] se proponen filtros cuya respuesta sea:

$$20 \log |H_p(e^{j\omega})| = \frac{-60m_p T}{T_r(\omega)} \Rightarrow |H_p(e^{j\omega})| = 10^{\frac{-3m_p T}{T_r(\omega)}} \quad p = 1, \dots, n \quad (5)$$

donde ω es la frecuencia normalizada ($0 \leq \omega = 2\pi fT \leq 1$), f es la frecuencia en hertz, T es el período de muestreo en segundos y $T_r(\omega)$ es una función que especifica los tiempos de reverberación deseados versus la frecuencia.

Estos filtros desplazan los polos del círculo unitario a una ubicación definida por $T_r(\omega)$, y en consecuencia introducen una “coloración” no deseada. Para evitar esto se utiliza un filtro correctivo $T(z)$ de la forma:

$$|T(e^{j\omega})|^2 = \frac{1}{T_r(\omega)} \quad (6)$$

2.5 Unión de las Reflexiones Tempranas y Tardías. Construcción de la RIR.

Las reflexiones tempranas (especulares y difusas) y la cola de reverberación deben ser empalmadas de manera que la transición sea continua e imperceptible al oyente.

Según Stewart y Sandler [13]: “el punto de transición será el punto más temprano en el tiempo donde la densidad de las reflexiones ha alcanzado un umbral perceptual de manera que las reflexiones individuales no pueden ser distinguibles”.

Basándose en análisis estadísticos de la reverberación, se puede establecer que el tiempo a partir del cual la RIR se asemeja a un proceso estocástico es:

$$t_m \approx \sqrt{V} \quad (7)$$

conocido como *mixing time* dado en milisegundos, donde V es el volumen del recinto en m^3 .

Utilizamos este tiempo como referencia para empalmar ambos tipos de reflexiones. Con este valor, se calculó la distancia r que recorre un rayo en ese tiempo viajando a la velocidad del sonido. Se generaron todas las fuentes virtuales cuya distancia al receptor sea igual o menor a r y se estableció como $1/r$ la amplitud inicial de las reflexiones tardías (una onda sonora se atenúa proporcionalmente a la inversa de la distancia que recorre). Además, se descartó la primera porción de la salida de la FDN debido a que no es suficientemente aleatoria. Para determinar cuánto descartar se utilizó la curtosis. La curtosis es un índice estadístico que refleja qué tanto se aleja un proceso de uno gaussiano.

3 Resultados

Se realizaron mediciones de RIR en una sala sencilla de un volumen aproximado de $117,26 \text{ m}^3$. Se ubicó a la fuente y al receptor en posiciones arbitrarias. En la Tabla 1 se muestran resultados preliminares de tiempos de reverberación obtenidos de la RIR simulada y la RIR medida, junto al porcentaje de error. Están dados para 1 kHz con ancho de banda de 1/3 de octava. Los parámetros son los recomendados por la norma ISO 3382:1997 [14].

Para la simulación se utilizó un orden de fuentes-imágenes de 3, y los valores de los parámetros para los filtros de difusión y para la FDN son los indicados anteriormente.

Tabla 1. Comparación de los tiempos de reverberación a 1 kHz con ancho de banda de 1/3 de octava

	Simulación	Medición	Error [%]
EDT [s]	1.941	2.176	1.166
T10 [s]	2.107	2.019	0.190
T20 [s]	2.098	2.196	0.199
T30 [s]	2.092	2.130	0.032

4 Conclusiones

Se ha desarrollado un modelo para construir una RIR con el objeto de que sea eficaz para ejecutarse en tiempo real. Para alcanzar este fin, se utilizaron evidencias psicoacústicas que permitieron introducir simplificaciones a los algoritmos de manera que resultan perceptualmente irrelevantes. Si bien los resultados preliminares son alentadores, es necesario realizar más validaciones objetivas y subjetivas.

En este momento se están llevando a cabo numerosas mediciones de recintos de formas arbitrarias y de características acústicas diversas, que serán comparadas con las RIR simuladas según los parámetros recomendados por la norma ISO 3382: 1997.

Referencias

1. Kleiner, M., Dalenback, B.I., Svensson, P.: Auralization-an overview. *J. Audio. Eng. Soc.* 41(11), 861–875 (1993)
2. Blauert J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA. The MIT Press. (1997)
3. Ramos, O.A., Jasá, V., Gorgas, J., Céspedes Daza, D.: Simulación binaural de espacios acústicos. In: *V Congreso Iberoamericano de Acústica*, paper ID: A090. Santiago de Chile, Chile (2006)

4. Tsingos, N., Gascuel, J.-D.: Acoustic simulation using hierarchical time-varying radiant exchanges. Reporte técnico no publicado (1997)
5. Céspedes Daza, D., Jasá, V., Ramos O.A.: Room acoustics simulation model using the image source method. In: XI Reunión de Trabajo en Procesamiento de la Información y Control. Río Cuarto, Argentina. (2005)
6. Borish, J.: Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.* 75(6), 1827–1836 (1984)
7. Schroeder, M.R.: Natural-sounding artificial reverberation. *J. Audio. Eng. Soc.* 10(3), 219–223 (1962)
8. Pellegrini, R.S.: A Virtual Reference Listening Room as an Application of Auditory Virtual Environments. PhD thesis, Ruhr-University Bochum (2001)
9. Jot, J.-M.: An analysis/synthesis approach to real-time artificial reverberation. In: *Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. II.221–II.224. San Francisco, USA. IEEE Press, New York (1992)
10. Jot, J.-M., Chaigne, A.: Digital delay networks for designing artificial reverberators. In: *90th AES Conv*, preprint 3030 (E-2) (1991)
11. Stautner, J., Puckette, M.: Designing multi-channel reverberators. *Computer Music Journal.* 6(1), 52–65 (1982)
12. Frenette, J.: Reducing artificial reverberation algorithm requirements using time-varying feedback networks. MS thesis, Music Eng. Tech. University of Miami (2000).
13. Stewart, R., Sandler, M.: Statistical measures of early reflections of room impulse responses. In: *10th Int. Conference on Digital Audio Effects (DAFx-07)*. Bordeaux, Francia (2007)
14. ISO 3382 (E): Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters (1997).