

Sistema de realidad acústica virtual en tiempo real: AVRS

F. C. Tommasini^{a,b}, O. A. Ramos^{b,c}, S. P. Ferreyra^b, & R. M. Guido^b

^a*Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Córdoba, Argentina*
fabian@tommasini.com.ar

^b*Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Córdoba, Argentina*

^c*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina*

RESUMEN: En este trabajo se presenta un sistema experimental dinámico e interactivo de realidad acústica virtual. El sistema, llamado AVRS, es de código abierto, flexible, escalable y modular. Funciona en una computadora estándar y se ejecuta sobre un sistema operativo de *hard real-time* basado en un kernel Linux. Permite simular un entorno auditivo virtual de forma y dimensiones arbitrarias con el propósito de provocar en el usuario la sensación de inmersión en dicho entorno. Para lograrlo, se realizaron simplificaciones a modelos físico-matemáticos basados en evidencias perceptuales. Se utilizó un modelo híbrido para sintetizar la respuesta impulsiva del recinto (room impulse response, RIR) basado en: (a) el método de la fuente-imagen para el sonido directo y las reflexiones tempranas, y (b) una red de retardo realimentada (feedback delay network, FDN) para las reflexiones tardías y la difusión. Además, se modeló el oyente mediante filtros digitales generados a partir de las funciones de transferencia de la cabeza (head-related transfer functions, HRTFs). La auralización de una señal anecoica se logra mediante un módulo de convolución de baja latencia en tiempo real. A partir de una simulación de un recinto de dimensiones pequeñas se evaluó el rendimiento del AVRS mediante mediciones de los tiempos de cálculo de los diferentes módulos y de la latencia total del sistema. Los resultados obtenidos fueron menores a los umbrales perceptualmente detectables según diferentes autores.

KEYWORDS: Acústica virtual, Simulación acústica en tiempo real, Método de la fuente-imagen, Red de retardo realimentada, HRTF.

1. INTRODUCCIÓN

Kleiner et al. [1], en 1993, señalaron que si mediante algún medio de reproducción se aplican a los tímpanos de un oyente las señales biológicamente correctas, será posible estimular en el mismo la sensación de inmersión en el entorno modelado. Este proceso, denominado *auralización*, es la base que sustenta a la realidad acústica virtual (acoustic virtual reality, AVR) y es análogo al término *visualización*, utilizado para la simulación de imágenes.

En un principio, los sistemas de auralización fueron desarrollados como módulos de programas de acústica de salas, que procesaban el sonido en modo *off-line* y luego lo reproducían [2]. No era posible realizar una simulación acústica en tiempo real por el gran poder de cómputo necesario para su cálculo. Hoy en día, aún con el incremento de la potencia de las computadoras personales, todavía no se puede lograr una simulación exacta de un entorno auditivo virtual (auditory virtual environment, AVE) en tiempo real, sino sólo es posible realizar simulaciones plausibles [3, 4]. Este abordaje procura simplificar el abordaje físico-acústico (que busca conseguir una reproducción auténtica) apelando a evidencias perceptuales, teniendo en cuenta aquellas características que son relevantes para lograr la sensación de inmersión en un AVE.

La sensación de inmersión en el entorno generado es un aspecto clave en los sistemas de realidad virtual. Dos requisitos son fundamentales para lograrla: continuidad y capacidad de respuesta. El primero se relaciona con la velocidad de cálculo para actualizar el escenario acústico, y el segundo, con la latencia. Si la velocidad no es la adecuada y/o la latencia es excesiva, el oyente percibirá una imagen auditiva discontinua y retrasada, respectivamente.

Por ello, los sistemas de realidad acústica virtual deben responder a eventos del entorno dentro de restricciones precisas de tiempo. Los sistemas operativos *hard real-time* aseguran satisfacer estas restricciones.

Desde hace algunos años, en el CINTRA de la Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina, se está desarrollando un sistema experimental de realidad acústica virtual en tiempo real dinámico e interactivo basado en una arquitectura de computadora de propósito general [5, 6, 7, 8]. Desde los primeros prototipos, en 2007, el sistema ha ido evolucionando, y se fueron incluyendo mejoras y cambios para optimizar su funcionamiento y eficiencia. Actualmente, permite simular campos sonoros de recintos con formas arbitrarias, limitado a una fuente sonora y a un oyente, el cual puede moverse libremente en el entorno virtual. El sistema, llamado AVRS, es gratuito y de código abierto¹, y combina flexibilidad, escalabilidad, modularidad, con la estabilidad provista por un sistema operativo *hard real-time*.

En este trabajo se presenta al sistema AVRS en su versión 0.5, que es la versión estable más reciente, y se muestran los resultados de rendimiento de una simulación realizada para un recinto de dimensiones pequeñas.

2. EL SISTEMA

El sistema AVRS está basado totalmente en software y escrito principalmente en lenguaje C++ (con algunos módulos en C), y se ejecuta en una computadora personal estándar. No requiere de ningún tipo de hardware específico, como por ejemplo, la utilización de un procesador digital de

¹avrsystem.sf.net

señales (digital signal processor, DSP) para procesamiento de audio. Sólo es necesario, de manera opcional, un sensor de movimientos de cabeza para proveer la realimentación de los movimientos del usuario. También brinda la posibilidad de cargar un archivo con movimientos y desplazamientos pre-grabados.

2.1 Ejecución en hard real-time

El correcto funcionamiento de un sistema de realidad acústica virtual no sólo depende de la precisión de los modelos sino también del momento en que los resultados son producidos. Los sistemas operativos *hard real-time* aseguran satisfacer estas restricciones de tiempo. Las tareas que pueden ejecutar tienen un tiempo límite o *deadline* para completarse. Estas tareas se clasifican en: tareas *hard real-time* (HRT) y tareas *soft real-time* (SRT) [9]. Una tarea HRT debe tener un cumplimiento estricto del *deadline*. Si lo supera podría causar serios problemas. En el caso de un AVE, escuchar “clics” o discontinuidades en la reproducción del sonido, por ejemplo. En cambio, en una tarea SRT no es necesario un cumplimiento estricto del *deadline*, pero sí es deseable por cuestiones de rendimiento. Si el *deadline* es superado, no causaría problemas graves. Por ejemplo, el sensado de los movimientos de cabeza del usuario.

El AVRS se ejecuta sobre un sistema operativo *hard real-time* basado en RTAI (Real-Time Application Interface) [10]. RTAI² es una extensión de *hard real-time* de código abierto para el kernel Linux. Estudios comparativos [11] demuestran que RTAI logra un mejor rendimiento (menor delay y menor jitter) que otros sistemas operativos *hard real-time* de código abierto, e inclusive que alternativas privativas-comerciales. La latencia de Linux-RTAI es menor a 10 μ s para una computadora estándar.

Durante su ejecución en tiempo real, el AVRS tiene un período de aproximadamente 11.6 ms (512 muestras a 44.1 kHz), lo que constituye una tasa de actualización de aproximadamente 86 Hz. Se sigue una estrategia de ejecución en paralelo de múltiples tareas, optimizando el uso del tiempo en los múltiples núcleos de las actuales CPUs.

2.2 Arquitectura del sistema

El software del sistema AVRS se puede dividir en un módulo de configuración, en procesos SRT y en procesos HRT. El módulo de configuración, desarrollado en MATLAB (The MathWorks, Inc.), fue creado para cargar la geometría del recinto y los parámetros del sistema de manera más simple. La configuración resultante se guarda en un archivo que es utilizado por el AVRS para parametrizar las simulación.

Una vez que se ejecuta el AVRS, los procesos SRT controlan la entrada del usuario, capturan los datos que envía el sensor de movimientos de cabeza y calculan una parte de la respuesta impulsiva binaural (binaural impulse response, BIR) del entorno virtual. Los procesos HRT realizan el cálculo de los modelos y el filtrado correspondiente para sintetizar la BIR, para luego realizar la convolución de baja latencia con una señal anecoica de entrada. El resultado es reproducido mediante auriculares.

²www.rtai.org

2.3 Hardware

Se utiliza una computadora personal basada en una arquitectura de propósito general, unos auriculares Sennheiser HD 570 y un sensor óptico de movimientos de cabeza de desarrollo propio [8, 12]. Dicha computadora personal consta de un procesador AMD FX-8320 de 8 núcleos de 3.5 GHz, 16 GB de memoria RAM DDR3 de 1600 MHz y una placa de sonido Intel HDA. Cualquier componente de hardware puede ser reemplazado y actualizado sin afectar al resto, por ejemplo, por otro sensor con menor latencia y/o más precisión.

3. IMPLEMENTACIÓN

3.1 Modelado del recinto

Cuando una onda sonora toma contacto con las superficies y objetos del entorno se desencadenan una serie de fenómenos físico-acústicos que modifican la dirección, espectro y magnitud de la onda. Las ondas sonoras, además, son alteradas por el medio de propagación (aire). Para lograr exactitud, el modelado del campo sonoro de un recinto se debería resolver mediante las ecuaciones de la onda en el espacio y en el tiempo (ecuaciones de Helmholtz), pero sólo las aproximaciones introducidas por los métodos basados en la teoría geométrica (acústica geométrica) permiten simular la respuesta impulsiva del recinto (room impulse response, RIR) en tiempo real.

Actualmente, los modelos determinísticos que se utilizan están basados en el método de la fuente-imagen (image-source method, ISM) [13, 14]. Este método provee una buena aproximación en recintos con superficies de baja absorción, pero desafortunadamente falla al simular grandes escenarios, ya que el costo computacional crece exponencialmente junto con la cantidad de reflexiones a simular [3]. Por ello, es habitual utilizar un modelado híbrido. El sonido directo y las reflexiones tempranas, que son importantes determinarlos con precisión desde el punto de vista perceptual para que el usuario pueda localizar una fuente sonora, se simulan con el ISM; mientras que la difusión y la cola de reverberación, que no son necesarias determinarlas con la misma precisión, con otro método de menor costo computacional. En este caso se utilizó una red de retardo realimentada (feedback delay network, FDN), que es una estructura de filtros recursivos.

Método de la fuente-imagen: El método de la fuente-imagen (ISM) es un método basado en rayos. Los rayos sonoros reflejados en una superficie son reemplazados por rayos directos provenientes de fuentes imágenes espejo reflejadas de la fuente real [13]. Para determinar una reflexión, una fuente imagen o fuente virtual (virtual source, VS) se encuentra equidistante de la fuente real según la normal a la superficie que le da origen.

Para encontrar todas los caminos de reflexión desde la fuente sonora a el oyente, el sonido de la fuente es reflejado contra todas las superficies del recinto. También hay imágenes de orden superior, las cuales representan sonidos que se reflejan primero contra una superficie y luego contra otra, y así sucesivamente. De esta manera, el campo sonoro total en la posición del oyente está conformado por el sonido directo, originado por la fuente real (recorre el mínimo camino entre la fuente y el oyente), y por el conjunto de reflexiones originadas en las VSs. Luego de encontrar todas las VSs, se debe realizar un chequeo de audibilidad (también llamado chequeo de visibilidad). Esto indica si cada VS es audible (o visible) para el oyente [5].

Es importante notar que la ubicación de las VSs no dependen de la posición del oyente. Sólo la audibilidad de cada fuente virtual puede cambiar cuando el oyente se mueve. Pero el

algoritmo de chequeo de la audibilidad es muy demandante computacionalmente, ya que requiere una gran cantidad de cálculos de intersecciones en superficies y de caminos de reflexión. Por ello, si una VS pasó el chequeo de audibilidad para la posición actual del oyente, el AVRS almacena todos los datos de dicha fuente audible en un vector dinámico para su utilización futura.

Red de retardo realimentada: El campo sonoro de las reflexiones tardías o cola de reverberación es usualmente considerado difuso y se asemeja a un ruido aleatorio que se atenúa exponencialmente con el tiempo. Una red de retardo realimentada (FDN) de ocho líneas [15] permite modelar la difusión y la cola de reverberación [5]. Tiene la característica de lograr solapamiento (o densidad) suficiente de los modos acústicos en el dominio de la frecuencia y de las reflexiones en el dominio del tiempo [16, 17]. Una ventaja es que la respuesta impulsiva resultante tiene una densidad de reflexiones creciente como función del tiempo, tal es el caso de las respuestas de los recintos reales [18].

El tiempo de decaimiento (tiempo de reverberación) de una respuesta de la FDN puede ser definida mediante filtros dependientes de la frecuencia en el contexto de las líneas de retardo. Para simular la difusión se atenuó con una función exponencial la primera parte de la respuesta de la FDN [8]. La salida de la FDN se calcula por única vez al comenzar la simulación, ya que se asume que la reverberación se mantiene constante mientras no cambie la geometría del recinto.

Síntesis del método híbrido: Cuando la RIR es modelada con un método híbrido, se requiere establecer un tiempo de transición específico entre las reflexiones tempranas y la cola de reverberación. Dicha transición debe ser continua e imperceptible al oyente. Este punto de transición determina la proporción del tamaño de la RIR que será simulado con un método y con el otro. Consecuentemente, esto tiene grandes influencias en la carga computacional resultante.

El sistema AVRS sugiere el tiempo de transición, el cual es calculado a partir de la fórmula del *mixing time* [17]: $t_m \approx \sqrt{V}$, donde V es el volumen del recinto en m^3 , y t_m es el tiempo expresado en ms. Según los autores, esta fórmula establece el tiempo a partir del cual la RIR se asemeja a un proceso estocástico. Pero también existe la posibilidad de modificar este valor, por otro determinado por algún método alternativo. Por ejemplo, algunos autores sugieren diferentes valores [15, 19, 20], o bien puede ser calculado a partir de una RIR medida [21, 22].

Con el valor del tiempo de transición t_m se calcula la distancia r que recorrerá una onda sonora en dicho tiempo: $r = c \cdot t_m / 1000$, donde c es la velocidad del sonido. Ésta es la distancia máxima para la cual se calcularán VSs. La amplitud inicial de la cola de reverberación se estimó con la ley $1/r$.

3.2 Modelado del oyente

Las transformaciones de las ondas sonoras producidas por el torso, los hombros y especialmente por la cabeza y los pabellones auriculares se encuentran presentes en las respuestas impulsivas de cabeza (head-related impulse responses, HRIRs). Aplicando la transformada de Fourier a las HRIRs se obtienen las funciones de transferencia de cabeza (head-related transfer functions, HRTFs). Estas HRTFs son determinadas para diferentes posiciones discretas de la fuente ubicada sobre una superficie esférica cuyo centro es ocupado por el oyente.

Utilizando la base de datos de HRIRs del CIPIC [23], las HRTFs fueron modeladas mediante filtros IIR de orden 20 por el método iterativo de Steiglitz-McBride [24]. Con tan sólo 5

iteraciones se lograron aproximaciones, tanto para la magnitud como para la fase, perceptualmente aceptables [25].

Se ha demostrado que una persona adulta puede distinguir entre dos fuentes sonoras separadas entre 1° y 2° en el frente [26, 27] y entre 3° a 9° en el hemisferio frontal del plano medio [28, 27]. Pero en la práctica, es muy difícil medir las HRTFs con esa resolución espacial. Sólo se miden para algunos ángulos discretos de azimut y elevación. En un AVE, para poder recrear una verídica percepción continua del espacio circundante se debe realizar algún tipo de interpolación espacial a la base de datos de HRTFs. Langendijk y Bronkhorst [29] realizaron pruebas psicoacústicas y determinaron que aplicando interpolación bilineal a HRTFs medidas a una resolución espacial entre 6° y 15° no se introducen claves audibles importantes y se puede localizar la fuente sonora con igual precisión, independientemente si la HRTF es medida o interpolada.

En base a estas evidencias, se realizó la interpolación espacial de las HRIRs de la base de datos del CIPIC a valores de azimut distanciados 1.25° entre sí, en el rango -80° a $+80^\circ$, y a valores de elevación distanciados 2.8125° entre sí, en el rango -45° y $+230.625^\circ$. Dando como resultado un total de 12771 posiciones (originalmente se contaba con 1250) para cada oído, con 129 posiciones en azimut y 99 posiciones en elevación [8].

3.3 Filtrado de las reflexiones tempranas y el sonido directo

En un sistema dinámico e interactivo, cada vez que el oyente realiza un movimiento por encima de un umbral (determinado en 5 cm para la traslación y 2.5° para la rotación, para evitar cálculos innecesarios o espurios) e interactúa con el entorno, implica recalcular la primera porción de la BIR. Tanto el sonido directo como las reflexiones tempranas pasan a través de varios filtros: directividad, absorción de los materiales de las superficies (sólo las reflexiones tempranas), atenuación del aire y distancia, y HRTF.

Filtros de directividad: Los filtros de directividad fueron diseñados de acuerdo al enfoque de filtrado direccional [18]. El sistema obtiene de una base de datos el filtro FIR de la fuente según la orientación de la reflexión correspondiente. Hasta ahora sólo se realizaron simulaciones con una fuente de sonido omnidireccional.

Filtros de absorción de materiales: Cuando las ondas sonoras interactúan con los materiales de las superficies, no toda la energía es reflejada. Por acción de los materiales, la onda reflejada tiene un contenido espectral diferente a la onda incidente. El cociente entre la energía que se refleja y la energía incidente se conoce como coeficiente de reflexión β . Es común disponer de los coeficientes de absorción α , independientes del ángulo de incidencia, para frecuencias discretas entre 125 y 4000 Hz espaciadas por bandas de una octava, donde $\alpha(\omega) = 1 - |\beta(j\omega)|^2$.

Lehmann y Johansson [30] proponen utilizar los coeficientes de reflexión β negativos, indicando que se obtienen resultados más precisos en la simulación. A partir de los coeficientes de reflexión negativos se diseñaron filtros IIR de cuarto orden para modelar la absorción de los materiales [18]. Éstos fueron interpolados linealmente para frecuencias intermedias y extrapolados para frecuencias superiores a 4000 Hz. Los filtros de absorción de materiales actúan sólo para las reflexiones tempranas, no para el sonido directo. La cantidad de estos filtros a aplicar a cada reflexión dependerá del orden de la VS (uno por cada orden o superficie).

Atenuación de aire y de distancia: Cuando una onda sonora viaja a través de un medio de transmisión (aire en este caso), éste absorbe parte de su energía. El efecto de la atenuación por la distancia recorrida se logra multiplicando la señal por la inversa de tal distancia.

Filtros HRTF: Hasta aquí se manipuló una señal monoaural, pero a partir del filtrado por HRTF se genera una señal binaural. Son un par de filtros dependientes de la dirección de incidencia del sonido en el oyente. Corresponden al oído izquierdo y al oído derecho, y luego se agrega el retardo correspondiente a la diferencia de tiempo interaural (interaural time difference, ITD). Se aplican a cada una de las reflexiones tempranas y al sonido directo. Realizando la síntesis de las salidas de estos filtros, se obtiene una señal binaural que contiene toda la información direccional. Se utilizaron los filtros IIR diseñados por el método de Steiglitz-McBride [25].

3.4 Convolución de baja latencia

La convolución de la señal anecoica de entrada con cada canal de la BIR sintetizada requiere de una gran potencia computacional. Para disminuir tanto el tiempo de cálculo como la latencia se utiliza el método de convolución por particiones no uniformes [31]. Este algoritmo realiza un acondicionamiento especial en bloques de toda la BIR y procesa cada bloque individualmente en el dominio de la frecuencia. Esta convolución sólo agrega un retardo de $2N$, por lo que es recomendable usar bloques lo más pequeños posible ($N = 256$ muestras en el AVRS). El algoritmo está basado en el motor de convolución del sistema SSR [32].

4. RESULTADOS

4.1 Caso de estudio

Para evaluar el rendimiento del AVRS, se simuló un aula de la Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina. El recinto tiene un volumen de aproximadamente 117 m^3 . Las dimensiones son 5.74 m de ancho, 6.93 m de largo y 3.03 m de alto.

Los parámetros generales para la simulación fueron: VSs de hasta 2° orden para el ISM, 2.5 s de largo de la BIR (≈ 110000 muestras a 44.1 kHz), y 25 ms de tiempo de transición.

4.2 Tiempos de cálculo

El tiempo que toma realizar una auralización de un sonido anecoico es un indicador del rendimiento del sistema. Depende de varios factores, tales como el recinto (dimensiones y complejidad geométrica), directividad de la fuente sonora, modelado del oyente, método para generación de la respuesta impulsiva, orden de los filtros, algoritmo de convolución de baja latencia, organización de las tareas en paralelo, entre otros.

Los tiempos promedio (y la desviación estándar, SD) que tomó calcular cada elemento de la cadena de filtros para las reflexiones tempranas y el sonido directo, se muestran en la Tabla 1.

La síntesis del método híbrido tomó $97.41 \mu\text{s}$ (SD = $2.18 \mu\text{s}$). Para completar el cálculo de las reflexiones tempranas mediante el ISM (cálculo y validación de 28 VSs audibles) se logró un tiempo promedio de 4.2 ms (SD = 0.05 ms). La auralización completa de un sonido anecoico tomó alrededor de 7.4 ms (SD = 0.16 ms).

Tabla 1: Promedio (y desviación estándar) de los tiempos de cálculo para la cadena de filtros

	Tiempos [μ s]
Filtro de directividad	0.80 (0.16)
Filtros de absorción de materiales (por superficie)	7.99 (0.24)
Atenuación de aire y distancia	0.40 (0.14)
Filtros HRTF (ambos oídos)	91.70 (2.95)
Acumulación en el buffer	0.97 (0.21)

4.3 Latencia total del sistema

La latencia es otro indicador de rendimiento. La latencia total del sistema (total system latency, TSL) se define como el tiempo transcurrido desde la transducción de un evento, tal como un movimiento de cabeza, hasta que las consecuencias de dicho evento causan los cambios necesarios en el entorno virtual, tal como la actualización de los resultados de la auralización. Al tiempo de latencia contribuyen varios componentes individuales, que incluyen el dispositivo de sensado de movimientos de cabeza, procesamiento de las señales y algoritmos de auralización y reproducción del sonido. También incluye las líneas de comunicación, intercambio de datos y controladores de los dispositivos de entrada/salida.

Para obtener un patrón de movimiento que sea repetible, el sensor de movimientos de cabeza fue montado sobre un disco de acrílico. Dicho disco fue mecanizado con máquinas herramientas de control numérico de alta precisión y cuenta con orificios a lo largo de su borde exterior. Se eligió uno de los orificios del disco como punto de referencia (azimut = 0°) y el resto de los orificios fueron obstruidos. Dicho valor fue usado por el software para comparar con los datos que posteriormente fueron enviados por el sensor. Cuando el disco pasa por el punto, un optoacoplador emite una señal escalón (tiempo T_0), y cuando la posición dada por el sensor coincide con la de referencia, el AVRS actualiza el escenario acústico virtual y envía una onda cuadrada³ a la placa de sonido (en lugar de enviar el sonido auralizado) (tiempo T_1). Tanto la señal generada por el optoacoplador como la señal de la placa de sonido fueron grabadas en un osciloscopio digital (Tektronix TDS210). La diferencia de ambos tiempos (T_1 y T_0) es la TSL.

La media para 30 mediciones fue de 46.13 ms, con una desviación estándar de 17.09 ms. El valor mínimo fue 23 ms y el valor máximo fue 77 ms. Estos valores incluyen mediciones de la latencia del sensor óptico, que actualiza la información de posición y orientación de la cabeza del usuario en aproximadamente 26 ms. Es posible reducir el tiempo de latencia del sistema recurriendo a otros sistemas comerciales de sensado que poseen latencias menores (~ 5 – 10 ms), por ejemplo, el Polhemus Patriot o el Xsens MTi-10.

5. DISCUSIÓN Y CONCLUSIONES

Como se mencionó anteriormente, un parámetro crítico en el diseño de AVEs es el máximo tiempo de latencia que se puede permitir (umbral detectable). No hay muchos estudios que indaguen sobre la mínima latencia detectable en un sistema de realidad acústica virtual. Wenzel [33] dedujo esta latencia partir de resultados de pruebas psicoacústicas sobre el mínimo ángulo audible en

³Dado que no es posible reproducir una salida de corriente continua en la placa de sonido.

movimiento (minimum audible movement angle, MAMA) [28]. Este valor indica el ángulo que una fuente en movimiento tiene que alcanzar para ser detectada como una fuente diferente a una estacionaria. Se puede inferir que los valores de MAMAs son igualmente válidos para un oyente que gira su cabeza y una fuente virtual estática. Los resultados arrojados por el estudio de Wenzel indican que los umbrales de detección van desde 59 a 92 ms, dependiendo de la velocidad de movimiento de la cabeza (rápido y lento, respectivamente).

En los últimos años, algunos estudios [34, 35, 36] midieron la mínima latencia detectable en diferentes AVEs. Los umbrales promedio reportados varían desde 60 a 85 ms, con un valor mínimo en 53 ms obtenido por el estudio de Lindau [34].

Los resultados de rendimiento indicaron que el AVRS es capaz de realizar una auralización de un aula de dimensiones pequeñas en un tiempo promedio de 7.4 ms y el valor de la TSL promedio es de 46.13 ms. Estos valores resultan perceptualmente aceptables, de acuerdo a los estudios anteriormente mencionados.

En síntesis, el sistema AVRS de realidad acústica virtual en tiempo real está basado en arquitectura de computadoras de propósito general y desarrollado bajo un sistema operativo *hard real-time* (kernel Linux modificado con la extensión RTAI). Este sistema permite simular campos sonoros de recintos con formas arbitrarias y el oyente puede moverse libremente en el entorno virtual. Para lograrlo se aplicaron modelos físico-matemáticos simplificados, basados en evidencias perceptuales.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Kleiner, M., Dalenbäck, B.-I., & Svensson, U. P. Auralization-an overview. *J. Audio Eng. Soc.*, 41(11), 861–875 (1993).
- [2] Naylor, G. ODEON—another hybrid room acoustical model. *Appl. Acoust.*, 38(2-4), 131–143 (1993).
- [3] Lentz, T., Schröder, D., Vorländer, M., & Assenmacher, I. Virtual reality system with integrated sound field simulation and reproduction. *EURASIP J. Adv. Sig. Pr.*, 2007, 1–17 (2007).
- [4] Scarpaci, J. W., Colburn, H. S., & White, J. A. A system for real-time virtual auditory space. *Proceedings of the International Conference on Auditory Display (ICAD 2005)* (2005).
- [5] Tommasini, F. C. & Ramos, O. A. Modelo basado en evidencias perceptuales para simular respuestas impulsivas de recintos. *XXIV Conferencia Latinoamericana de Informática, CLEI 2008*. Santa Fe, Argentina (2008).
- [6] Tommasini, F. C., Ramos, O. A., & Ferreyra, S. Modelado eficiente y validación objetiva de la respuesta impulsiva de un recinto. *Mecánica Computacional*, XXVII, 443–455 (2008).
- [7] Tommasini, F. C. & Ramos, O. A. Auralización en tiempo real: implementación del modelo del oyente. *2º Congreso Internacional de Acústica UNTREF*. Buenos Aires, Argentina (2010).
- [8] Tommasini, F. C. Sistema de simulación acústica virtual en tiempo real. *Doctoral thesis, Universidad Nacional de Córdoba. Facultad de Ciencias Exactas, Físicas y Naturales, Córdoba, Argentina.* (2012).
- [9] Buttazzo, G. C. *Hard real-time computing systems: predictable scheduling algorithms and applications*. Springer-Verlag New York Inc (2004).
- [10] Mantegazza, P., Dozio, E. L., & Papacharalambous, S. RTAI: Real time application interface. *Linux J.*, 2000(72es), 10 (2000).
- [11] Barbalace, A., Luchetta, A., Manduchi, G., Moro, M., Soppelsa, A., & Taliercio, C. Performance comparison of VxWorks, linux, RTAI, and xenomai in a hard real-time application. *IEEE Transactions on Nuclear Science*, 55(1), 435–439 (2008).
- [12] Araneda, M., Tommasini, F. C., Cravero, G. A., Ferreyra, S., Novillo, D., & Ramos, O. A. Sensor de posición de cabeza para ser usado en un sistema de realidad acústica virtual en tiempo real. *Reporte técnico, CINTRA, UTN FRC, Córdoba, Argentina* (2011).

- [13] Allen, J. B. & Berkley, D. A. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, 65(4), 943–950 (1979).
- [14] Borish, J. Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.*, 75(6), 1827–1836 (1984).
- [15] Jot, J.-M. An analysis/synthesis approach to real-time artificial reverberation. volúmen 2, págs. 221–224 vol.2 (1992).
- [16] Schroeder, M. R. Natural sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3), 219–223 (1962).
- [17] Jot, J.-M., Cerveau, L., & Warusfel, O. Analysis and synthesis of room reverberation based on a statistical time-frequency model. *Audio Engineering Society Convention 103* (1997).
- [18] Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47(9), 675–705 (1999).
- [19] Garas, J. *Adaptive 3D sound systems*. Springer Netherlands (2000).
- [20] Blesser, B. A. An interdisciplinary synthesis of reverberation viewpoints. *J. Audio Eng. Soc.*, 49(10), 867–903 (2001).
- [21] Stewart, R. & Sandler, M. Statistical measures of early reflections of room impulse responses. *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, págs. 59–62 (2007).
- [22] Hidaka, T., Yamada, Y., & Nakagawa, T. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *J. Acoust. Soc. Am.*, 122(1), 326–332 (2007).
- [23] Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. The CIPIC HRTF database. *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, págs. 99–102. New Paltz, NY, USA (2001).
- [24] Steiglitz, K. & McBride, L. A technique for the identification of linear systems. *IEEE T. Automat. Contr.*, 10(4), 461–464 (1965).
- [25] Ramos, O. A., Araneda, M., & Tommasini, F. C. Diseño y evaluación de filtros binaurales. *Mecánica Computacional, XXVIII*, 137–148 (2009).
- [26] Mills, A. W. On the minimum audible angle. *J. Acoust. Soc. Am.*, 30(4), 237–246 (1958).
- [27] Blauert, J. *Spatial hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, MA (1997).
- [28] Perrott, D. R. & Saberi, K. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87, 1728–1731 (1990).
- [29] Langendijk, E. H. A. & Bronkhorst, A. W. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.*, 107(1), 528–537 (2000).
- [30] Lehmann, E. A. & Johansson, A. M. Prediction of energy decay in room impulse responses simulated with an image-source model. *J. Acoust. Soc. Am.*, 124(1), 269–277 (2008).
- [31] Müller-Tomfelde, C. Time-varying filter in non-uniform block convolution. *Proc. of the COST G-6 Conference on Digital Audio Effects*. Limerick, Ireland (2001).
- [32] Ahrens, J., Geier, M., & Spors, S. The SoundScape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. *AES Convention 124*, pág. 7330 (2008).
- [33] Wenzel, E. M. The impact of system latency on dynamic performance in virtual acoustic environments. *Proceedings of the 16th I International Congress of Acoustics and 135th Meeting of the Acoustical Society of America*, volúmen 135, pág. 180. Seattle, WA (1998).
- [34] Lindau, A. The perception of system latency in dynamic binaural synthesis. *Proc. of NAG/DAGA*, págs. 1063–1066. Rotterdam, The Netherlands (2009).
- [35] Mackensen, P. *Auditive Localization. Head movements, an additional cue in Localization*. Doctoral thesis, Technische Universität, Berlin (2004).
- [36] Yairi, S., Iwaya, Y., & Suzuki, Y. Investigation of system latency detection threshold of virtual auditory display. *Proc. of ICAD 2006-12th Meeting of the International Conference on Auditory Display*, págs. 217–222 (2006).