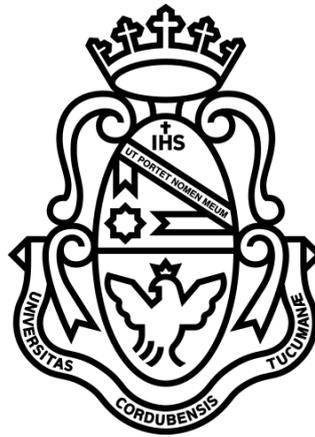


Universidad Nacional de Córdoba
Facultad de Ciencias Exactas, Físicas y Naturales



Tesis Doctoral

**Sistema de simulación acústica virtual
en tiempo real**

Fabián Carlos Tommasini

2012

Tesis Doctoral

Sistema de simulación acústica virtual en tiempo real

por

Ing. Fabián Carlos Tommasini

Director:

Ing. Oscar Alberto Ramos

Comisión Asesora:

Ing. Mario René Serra

Dr. Luis Rafael Canali

Tribunal Evaluador:

Dr. Ricardo A. Burdisso

Dr. Fernando Alberto Marengo Rodriguez

Dr. Enrique Adolfo Suárez Silva

Tesis presentada a la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba en cumplimiento parcial de los requerimientos para la obtención del grado académico de Doctor en Ciencias de la Ingeniería.

Córdoba, Argentina

Marzo de 2012



Doctorado en Ciencias de la Ingeniería
Escuela de Cuarto Nivel Académico
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de Córdoba
República Argentina

Avda. Vélez Sarsfield 1601 – Tel. 54 351 433 4404 - 5000 Córdoba

Acta Tribunal de Tesis de Doctorado

En la Ciudad de Córdoba, a los 30 días del mes de Marzo de 2012 se reúne el Tribunal de Tesis del Doctorado en Ciencias de la Ingeniería, del Ing. Fabián Carlos Tommasini, cuya tesis se denomina “Sistema de Simulación Acústica Virtual en Tiempo Real”.

El tribunal, integrado por el Dr. Ricardo A. Burdisso, el Dr. Fernando Alberto Marengo Rodriguez y el Dr. Enrique Adolfo Suárez Silva, después de haber analizado la tesis durante el período previo a su defensa y de presenciar su exposición oral, acuerda por unanimidad calificar la Tesis como:

APROBADA

A los efectos de dejar constancia se firma la presente acta.

A handwritten signature in black ink, appearing to read 'R. Burdisso', written over a horizontal dashed line.

Dr. Ricardo A. Burdisso
Virginia Polytechnic Institute
and State University

A handwritten signature in black ink, appearing to read 'F. Marengo', written over a horizontal dashed line.

Dr. Fernando A. Marengo Rodriguez
Universidad Nacional de Rosario

A handwritten signature in black ink, appearing to read 'E. Suárez Silva', written over a horizontal dashed line.

Dr. Enrique A. Suárez Silva
Universidad Austral de Chile

A Mercedes y a nuestro bebé Genaro

A mis padres y hermano

A mi abuela Marta y abuelo Ángel

Agradecimientos

Este trabajo de tesis no hubiera sido posible sin la colaboración de una gran cantidad de personas. Las palabras no me alcanzan para expresar la inmensa gratitud que siento hacia cada una de ellas, aunque intentaré manifestarla en estas líneas.

A mi director, Oscar Ramos, quien me ha brindado su apoyo incondicional y ha sido mi guía para la realización de este trabajo. Un gran investigador del cual he aprendido mucho más que lo que reflejan estas páginas, tanto a nivel profesional como personal. Por su confianza y su paciencia durante todos estos años. Por su generosidad y por sus consejos que me ayudaron a crecer.

A los miembros de la comisión asesora, Mario René Serra y Luis Canali, por el tiempo que me dedicaron y por sus valiosas opiniones y observaciones.

A los miembros del tribunal de tesis, Dr. Ricardo A. Burdisso, Dr. Fernando A. Marengo Rodriguez y Dr. Enrique A. Suárez Silva, por sus sugerencias y comentarios que ayudaron a enriquecer el trabajo.

A todos los integrantes del Centro de Investigación y Transferencia en Acústica (CINTRA) de la UTN FRC, los que están y los que estuvieron, por haberme dado un lugar, por su invaluable calidez humana y por su inmensa calidad profesional.

Quiero hacer mención especial a mis compañeros de línea de investigación:

Sebastián Ferreyra, Mariano Araneda y Agustín Cravero, por el trabajo compartido y sus valiosos aportes para esta tesis. También a David Novillo por su colaboración en la construcción de los dispositivos mecánicos necesarios.

A Aldo Ortíz, por los constructivos intercambios de ideas y su buena predisposición para brindarme ayuda cuando fuese necesario.

A Claudia Arias, por sus preciados comentarios y consejos que contribuyeron a mi formación en investigación científica.

A Juan Gorgas, quien me ayudó a abrir las puertas a este camino que estoy culminando.

A toda mi familia y amigos, por el aliento y la paciencia que tuvieron durante mis años de estudio.

Por último, quiero agradecer a mi pareja, Mercedes, por su respaldo constante, por su aliento y por el tiempo que me concedió para que pueda concluir esta nueva etapa.

A todos y a cada uno, les agradezco con el corazón...

Fabián C. Tommasini

Prefacio

Esta tesis se pudo realizar gracias al programa de becas doctorales internas del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET). Contiene los resultados de la investigación llevada a cabo en el Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET, de la Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN FRC), durante el período comprendido entre los años 2007 y 2011, bajo la dirección del Ing. Oscar A. Ramos, investigador del CONICET y docente de la UTN FRC.

Sistema de simulación acústica virtual en tiempo real

Resumen

Un sistema de realidad acústica virtual permite simular entornos acústicos cerrados con el propósito de producir en el usuario la sensación de inmersión en dicho entorno. Para lograr una convincente ilusión en tres dimensiones, el sistema debe ser capaz de responder a los cambios de escenarios que se producen por la acción del usuario en un tiempo y con una velocidad de actualización que garantice una continuidad perceptual aceptable. El objetivo de la presente tesis consistió en desarrollar un sistema experimental dinámico e interactivo de realidad acústica virtual, basado en arquitectura de computadoras de propósito general y que fuera flexible, escalable, fácil de mantener y modular. Para alcanzar este objetivo se modelaron tres subsistemas: la fuente sonora, el recinto y el medio de propagación, y el oyente. Se recurrió a simplificaciones tanto físicas como perceptuales para lograr modelos eficientes que sean factibles de operar en tiempo real y reproducir el escenario acústico mediante auriculares. Los resultados mostraron que los tiempos de ejecución y de latencia del sistema están por debajo de los umbrales de discriminación del ser humano.

A real-time virtual acoustic simulation system

Abstract

An acoustic virtual reality system enables to simulate closed acoustic environments in order to create in a user the feeling of immersion in that environment. To achieve a convincing three dimensional illusion, the system must be able to respond to changing scenarios, caused by user actions, at a time and update rate that ensure an acceptable perceptual continuity. The objective of this thesis was to develop a dynamic and interactive experimental virtual acoustic reality system, based on general purpose computer architecture. The system should be flexible, scalable, maintainable and modular. To achieve this objective three subsystems were modeled: the sound source, the room and the propagation medium, and the listener. Physical and perceptual simplifications were used to achieve efficient models that are feasible to operate in real-time and to reproduce the acoustic scenario through headphones. The results showed that the system latency and execution time are below the threshold that a human being is able to discriminate.

Sistema de simulação acústica virtual em tempo real

Resumo

Um sistema de realidade acústica virtual permite simular ambientes acústicos fechados a fim de produzir ao usuário a sensação de imersão nesse ambiente. Para alcançar uma ilusão convincente de três dimensões, o sistema deve ser capaz de responder às mudanças de cenários que são causados por uma ação do usuário, em um tempo e com uma taxa de atualização para garantir uma continuidade aceitável perceptual. O objetivo desta tese foi desenvolver um sistema experimental de realidade acústica virtual dinâmico e interativo, baseado em arquitetura de computadores de uso geral e que foi flexível, escalável, fácil de manter e modular. Para atingir este objetivo foram modelados três subsistemas: a fonte de som, a sala eo meio de propagação, eo ouvinte. Foi usado simplificações física e perceptual para alcançar modelos eficientes, que são viáveis para operar em tempo real e reproduzir o palco sonoro através de fones de ouvido. Os resultados mostraram que os tempos de execução e latência do sistema estão abaixo do limiar de discriminação dos seres humanos.

Índice general

Agradecimientos.....	vii
Prefacio.....	ix
Resumen.....	x
Abstract.....	xi
Resumo.....	xii
Índice general.....	xiii
Índice de figuras.....	xx
Índice de tablas.....	xxvii
Lista de abreviaturas.....	xxviii
Capítulo 1. Introducción.....	1
1.1 Subsistemas a modelar.....	5
1.2 Aplicaciones de los sistemas de simulación acústica.....	6
1.3 Objetivo y alcance de la tesis.....	7
1.4 Organización del informe de tesis.....	8
Capítulo 2. Fundamentos de acústica y psicoacústica.....	10
2.1 Propagación de las ondas sonoras.....	10

2.1.1 Velocidad del sonido.....	11
2.1.2 Ecuación de onda.....	12
2.1.3 Onda plana.....	13
2.1.4 Onda esférica.....	15
2.1.5 Energía y densidad de energía.....	16
2.1.6 Nivel de presión sonora.....	17
2.1.7 Intensidad sonora.....	18
2.2 Leyes de la reflexión sonora.....	18
2.2.1 Factor de reflexión y coeficiente de absorción.....	19
2.2.2 Impedancia sonora de la superficie.....	20
2.2.3 Reflexión del sonido con incidencia normal.....	20
2.2.4 Reflexión del sonido con incidencia oblicua.....	21
2.3 Difracción, dispersión y reflexión difusa.....	23
2.4 Campo sonoro difuso.....	26
2.5 Reverberación.....	26
2.6 Psicoacústica: audición espacial.....	28
2.6.1 Localización de fuentes sonoras.....	29
2.6.2 Claves involucradas en la localización.....	29
2.6.3 Localización en campo libre.....	31
2.6.4 Lateralización.....	32
2.6.5 Movimientos de cabeza en la localización sonora.....	33
2.6.6 Audición en ambientes reverberantes: efecto precedente.....	34
Capítulo 3. Fuente sonora.....	37
3.1 Señales de entrada al AVRS.....	37
3.2 Directividad de una fuente sonora.....	38

3.2.1 Instrumentos musicales.....	40
3.2.2 Voz humana.....	40
3.3 Modelado de la directividad.....	41
3.4 Fuente sonora omnidireccional.....	42
Capítulo 4. Recinto.....	45
4.1 Modelado de la acústica de recintos.....	45
4.2 Geometría del recinto.....	48
4.3 Respuesta impulsiva del recinto.....	51
4.4 Reflexiones especulares.....	54
4.4.1 Método de la fuente-imagen.....	54
4.4.2 Cálculo de la posición de una VS.....	56
4.4.3 Generación de las reflexiones tempranas especulares.....	57
4.4.4 Distribución temporal del campo sonoro en el oyente.....	62
4.5 Reflexiones difusas.....	64
4.5.1 Modelo de reflexiones difusas de Pellegrini.....	65
4.6 Caracterización del medio de transmisión.....	67
4.7 Caracterización de los materiales.....	68
4.8 Reflexiones tardías.....	71
4.8.1 Reverberador de Schroeder.....	72
4.8.2 Red de retardo realimentada.....	72
4.8.3 Matriz de realimentación y retardos.....	74
4.8.4 Ajuste a los tiempos de reverberación requeridos.....	75
4.8.5 Corrección tonal.....	76
4.8.6 Actualización de la FDN en el dominio del tiempo.....	77
4.9 Síntesis de la RIR.....	77

Capítulo 5. Oyente.....	80
5.1 Localización auditiva.....	80
5.2 HRIRs y HRTFs.....	85
5.3 Influencia del oído externo en las HRTFs.....	86
5.3.1 Canal auditivo.....	87
5.3.2 Orejas.....	87
5.4 Influencia de la cabeza en las HRTFs.....	91
5.5 Influencia del torso y hombros en las HRTFs.....	94
5.6 Abordajes para el modelado del oyente.....	95
5.6.1 Abordaje estructural.....	95
5.6.2 Abordaje empírico.....	95
5.6.3 Abordaje funcional.....	97
5.7 Criterios de evaluación del modelo del oyente.....	97
5.8 El método iterativo de Steiglitz-McBride.....	99
5.9 Interpolación de HRTFs.....	101
5.9.1 Interpolación bilineal.....	102
5.10 Personalización de HRTFs.....	104
5.10.1 Primeros intentos de personalización de HRTFs.....	105
5.10.2 Métodos actuales de personalización de HRTFs.....	106
Capítulo 6. Implementación del sistema.....	109
6.1 Reseña sobre sistemas de simulación acústica.....	109
6.2 Sistemas en tiempo real.....	111
6.2.1 Tareas de tiempo real.....	112
6.2.2 Planificación de las tareas de tiempo real.....	113

6.2.3 GNU/Linux como RTOS.....	114
6.3 Desarrollo del AVRS.....	117
6.3.1 Estructura de alto nivel.....	121
6.4 Ejecución en tiempo real.....	124
6.4.1 Actualización de los parámetros del sistema.....	128
6.4.2 Filtros para el sonido directo y las reflexiones tempranas	129
6.4.3 Generación de las reflexiones tardías.....	131
6.4.4 Convolución de baja latencia.....	133
6.5 Captura de los movimientos de cabeza.....	135
6.5.1 Cámara del wiimote.....	137
6.5.2 Beacon.....	138
6.5.3 Montaje del sensor de movimientos de cabeza.....	139
6.5.4 Cálculo de la posición y orientación.....	140
6.6 Sistemas de coordenadas en el AVRS.....	141
6.7 Software y hardware utilizados.....	143
6.7.1 Dependencias de software.....	143
6.7.2 Hardware.....	144
Capítulo 7. Resultados de un caso de estudio.....	146
7.1 Presentación del caso.....	146
7.2 Medición acústica del recinto.....	148
7.2.1 Fuente sonora.....	148
7.2.2 Señal de excitación.....	148
7.2.3 Equipamiento utilizado.....	149
7.2.4 Punto de medición.....	149
7.2.5 Resultados de la medición.....	151

7.3 Simulación acústica del recinto.....	152
7.3.1 Modelo de la fuente sonora.....	152
7.3.2 Modelo del recinto.....	153
7.3.3 Resultados de la simulación.....	155
7.4 Validación de la simulación acústica del recinto.....	156
7.5 Validación del modelo del oyente.....	162
7.5.1 Base de datos de HRTFs.....	163
7.5.2 Procedimiento.....	164
7.5.3 Errores en el plano horizontal.....	165
7.5.4 Errores en las ILDs.....	166
7.5.5 Errores en las ITDs.....	168
7.5.6 Errores en el plano medio.....	169
7.6 Validación del sensor de movimientos de cabeza.....	176
7.6.1 Errores en el plano horizontal.....	176
7.6.2 Errores en el plano medio.....	178
7.7 Rendimiento del AVRS.....	181
7.8 Medición de los tiempos de latencia.....	181
7.8.1 Latencia del sensor de movimientos de cabeza.....	183
7.8.2 Latencia total del AVRS.....	186
Capítulo 8. Conclusiones.....	189
8.1 Síntesis de la implementación y validación del sistema.....	189
8.2 Principales aportes y conclusión general.....	198
8.3 Trabajo a futuro.....	200
8.4 Publicaciones relacionadas a la tesis.....	202

Apéndice A. Parámetros acústicos de recintos.....	206
A.1 Parámetros relacionados a la reverberación.....	206
A.2 Parámetros relacionados a la energía.....	208
Apéndice B. Sistemas de referencia de la cabeza.....	211
B.1 Planos y ejes.....	211
B.2 Sistemas de coordenadas.....	212
B.2.1 Coordenadas polares referidas al eje longitudinal.....	212
B.2.2 Coordenadas polares referidas al eje interaural.....	213
Referencias.....	215

Índice de figuras

Capítulo 1

Figura 1.1 Diagrama de los tres subsistemas de un sistema de realidad acústica virtual.....	5
---	---

Capítulo 2

Figura 2.1 Frentes de onda de una onda plana.....	14
Figura 2.2 Frentes de onda de una onda esférica (vista en dos dimensiones).....	16
Figura 2.3 Reflexión del sonido con incidencia normal. Adaptado de Kuttruff (2009).....	21
Figura 2.4: Reflexión del sonido con incidencia oblicua. Adaptado de Kuttruff (2009).....	22
Figura 2.5 Difracción de una onda plana en el borde de un plano rígido. Extraído de Kuttruff (2009).....	24
Figura 2.6 Reflexión difusa en una superficie de estructura irregular. Extraído de Kuttruff (2009).....	25
Figura 2.7 (a) Representación esquemática del efecto precedente. Extraído de Arias y Ramos (2003). (b) Efecto precedente. Extraído de Arias (2009).....	36

Capítulo 3

Figura 3.1 Diagramas polares de directividad en el plano horizontal de un oboe para diferentes frecuencias. La dirección 0° corresponde al eje del instrumento. Extraído de Meyer (2009).....	39
Figura 3.2 Diagramas de directividad tridimensionales de un oboe para diferentes frecuencias. Extraído de Lentz (2007).....	39
Figura 3.3 Diagramas de directividad tridimensionales de la voz cantada para diferentes frecuencias. Adaptado de Vorländer (2007).	40
Figura 3.4 Esquema del filtrado direccional.....	42
Figura 3.5 Fuente sonora omnidireccional.....	43
Figura 3.6 Diagrama de directividad de la OSS, en campo libre.....	43
Figura 3.7 Respuesta impulsiva de la OSS y su magnitud espectral.....	44

Capítulo 4

Figura 4.1 Modelos para la acústica de recintos.....	47
Figura 4.2 Vista del modelo de un recinto con muchos detalles pequeños. Extraído de Vorländer (2007).....	49
Figura 4.3 Vista del modelo de un recinto simplificado desde el punto de vista acústico. Extraído de Vorländer (2007).....	50
Figura 4.4 Orden de generación de los vértices.....	50
Figura 4.5 Extracto de una RIR para visualizar el sonido directo, las reflexiones tempranas y las reflexiones tardías.....	52
Figura 4.6 La distancia recorrida y la dirección de arribo del rayo que se refleja en la pared son iguales a las del rayo emitido por la VS.....	55
Figura 4.7 Los rayos reales (que se reflejan en superficies) se reemplazan por los rayos que emiten las VSs.....	56
Figura 4.8 Cálculo de la posición de una VS.....	57

Figura 4.9 Determinación si el punto pertenece al polígono. En (a) el punto pertenece a la superficie evaluada; en (b) el punto no pertenece a la superficie.....	60
Figura 4.10 Prueba de audibilidad de orden superior.....	62
Figura 4.11 Las VSs contribuyen con un impulso discreto retardado y atenuado según la distancia al oyente.....	63
Figura 4.12 Reflexiones difusas.....	64
Figura 4.13 Reflexión difusa producida a partir de un impulso.....	65
Figura 4.14 Respuesta impulsiva del filtro $G(z)$ (sin sobremuestreo).....	67
Figura 4.15 Magnitud de los filtros de absorción de aire, sin tener en cuenta la atenuación por la distancia.....	68
Figura 4.16 Comportamiento de las ondas sonoras cuando inciden sobre una superficie.....	69
Figura 4.17 Coeficientes para un piso de carpeta de 1/8". Los puntos negros son los coeficientes de absorción α . Los puntos blancos, los coeficientes de reflexión R correspondientes. La línea continua se obtiene por interpolación lineal y extrapolación.....	70
Figura 4.18 FDN propuesta por Jot. Extraída de Frenette (2000).....	73

Capítulo 5

Figura 5.1 Ejemplo de la ILD (a) para baja frecuencia y (b) para alta frecuencia. Adaptado de Goldstein (2005).....	81
Figura 5.2 Ejemplo de la ITD.....	82
Figura 5.3 Cono de confusión.....	83
Figura 5.4 Tipos de movimientos de cabeza: rotación, tip y pivoteo.....	85
Figura 5.5 Respuesta impulsiva y magnitud de la función de transferencia del oído izquierdo (arriba) y del oído derecho (abajo) para una fuente sonora ubicada en	

azimut $+45^\circ$ y elevación 0°	86
Figura 5.6 Modos normales identificados por Shaw.....	87
Figura 5.7 Zonas de la oreja. Extraído de Shaw (1997).....	89
Figura 5.8 HRTFs para el plano medio. Extraído de Ramos, Tommasini, y Araneda (2010).....	90
Figura 5.9 Variación del espectro para el oído izquierdo, en función del ángulo de incidencia y el radio de la cabeza. Extraído de de Shaw (1997).....	92
Figura 5.10 HRTFs para el plano horizontal. Extraído de Ramos, Tommasini, y Araneda (2010).....	93
Figura 5.11 HRIRs para el plano horizontal. Extraído de Ramos, Tommasini, y Araneda (2010).....	94
Figura 5.12 Puntos a calcular en una interpolación bilineal	103

Capítulo 6

Figura 6.1: Estructura de capas de RTAI. Extraído de Barbalace et al. (2008)...	116
Figura 6.2 Cálculo y visualización de las VSs de un recinto.....	118
Figura 6.3 Visualización de la trayectoria de un rayo para una VS de segundos orden.....	119
Figura 6.4 Comparación objetiva de RIRs medidas y simuladas.....	120
Figura 6.5 Software de administración de pruebas psicoacústicas que permite evaluar modelos del oyente de diferentes grados de complejidad.....	121
Figura 6.6 Estructura de alto nivel del AVRS.....	122
Figura 6.7 Interfaz de configuración del AVRS.....	123
Figura 6.8 Diagrama esquemático de la arquitectura del AVRS.....	126
Figura 6.9 Combinación de las reflexiones tempranas con las reflexiones tardías.	132
Figura 6.10 Acondicionamiento de los bloques para la convolución particionada no	

uniforme. Adaptado de Lentz (2007).....	134
Figura 6.11 Diferentes vistas del wiimote. En la imagen de la derecha se puede ver la cámara del dispositivo.....	136
Figura 6.12 Diseño del beacon.....	137
Figura 6.13 Medidas de la ubicación de los LEDs en el beacon.....	138
Figura 6.14 Montaje del sensor de movimientos.....	138
Figura 6.15 Montaje final del AVRS.....	139

Capítulo 7

Figura 7.1 Render del recinto. a) Vista exterior, y b) vista interior.....	146
Figura 7.2 Puntos de medición. Para el caso de estudio se eligió el punto M5.....	149
Figura 7.3: Fotografías del aula al momento de la medición.....	150
Figura 7.4 RIR medida según norma ISO 3382.....	150
Figura 7.5 Posición de la fuente sonora (asterisco) y del oyente (triángulo).....	153
Figura 7.6 Distribución espacial de las VSs de primer y segundo orden.....	153
Figura 7.7 Trayectoria de los rayos para (a) VS de primer orden, y (b) VS de segundo orden.....	154
Figura 7.8 Respuesta impulsiva obtenida de la simulación.....	155
Figura 7.9 Gráficos de (a) EDT, (b) T10, (c) T20 y (d) T30 obtenidos de la RIR medida (línea continua) y de la RIR simulada (línea a trazos).....	156
Figura 7.10 Gráficos de (a) C50, (b) C80, (c) D50 y (d) Ts obtenidos de la RIR medida (línea continua) y de la RIR simulada (línea a trazos).....	157
Figura 7.11 Histograma de errores por frecuencia para (a) EDT, (b) T10, (c) T20 y (d) T30. Las jnd se muestran en línea a trazos.....	159
Figura 7.12 Histograma de errores por frecuencia para (a) C50, (b) C80, (c) D50 y (d) Ts. Las jnd se muestran en línea a trazos.....	160

Figura 7.13 Ubicación de los puntos de medición: (a) vista frontal, y (b) vista lateral. Extraído de Algazi et al. (2001).....	163
Figura 7.14 Magnitud espectral para una HRTF medida y estimada, para la posición 0° de azimut y -45° de elevación.....	164
Figura 7.15 Correlación cruzada entre las HRIR medidas y las modeladas del oído derecho para el plano horizontal.....	165
Figura 7.16 Error promedio de la diferencia entre las ILDs para HRTFs medidas y las ILD para HRTFs estimadas.....	166
Figura 7.17 Error medio entre los ITD medidos y los ITD aproximados (trazo negro). En puntos, los errores individuales.....	167
Figura 7.18 HRTFs de 35 sujetos.....	169
Figura 7.19 Correlación cruzada entre las HRIRs medidas y las HRIRs estimadas	170
Figura 7.20 Ventana Hann modificada aplicada a las HRIRs.....	172
Figura 7.21 Error en el modo 1 de Shaw.....	173
Figura 7.22 Errores de (a) presión sonora, y de (b) frecuencias centrales del primer notch.....	174
Figura 7.23 Sistema de referencia para el plano horizontal.....	175
Figura 7.24 Valores de referencia y valores medidos en el plano horizontal.....	176
Figura 7.25 Sistema de referencia para el plano medio.....	178
Figura 7.26 Valores de referencia y valores medidos en el plano medio.....	178
Figura 7.27 Eventos y etapas para medir la TSL.....	181
Figura 7.28 Configuración experimental para la medición de la latencia.....	182

Apéndice A

Figura A.1 Integración de Schroeder.....	205
--	-----

Apéndice B

Figura B.1 Planos y ejes de referencia de la cabeza. Adaptado de Scarpaci (2006).	210
Figura B.2 Coordenadas polares referidas al eje longitudinal.....	211
Figura B.3 Coordenadas referidas al eje interaural.....	212

Índice de tablas

Capítulo 6

Tabla 6.1 Eventos a considerar en un sistema de simulación acústica virtual en tiempo real.....	127
---	-----

Capítulo 7

Tabla 7.1 Equipamiento utilizado en la medición de la RIR.....	147
Tabla 7.2 Parámetros característicos de la RIR medida.....	150
Tabla 7.3 Parámetros característicos de la RIR simulada.....	154
Tabla 7.4 Errores calculados entre la RIR medida y la simulada, junto a la jnd correspondiente.....	157
Tabla 7.5 Tiempos de cálculo.....	179
Tabla 7.6 Valores de latencia del sensor de movimientos de cabeza.....	182
Tabla 7.7 Valores de latencia extremo a extremo.....	184

Lista de abreviaturas

ACPI	Advanced configuration and power interface (Interfaz avanzada de configuración y energía)
Adeos	Adaptive domain environment operating systems
ANN	Artificial neural network (Red neuronal artificial)
API	Application programming interface (Interfaz de programación de aplicaciones)
APM	Advanced power management (Administración avanzada de energía)
ASW	Ancho aparente de la fuente (Apparent source width)
BIOS	Basic input/output system (Sistema básico de entrada/salida)
BIR	Binaural impulse response (Respuesta impulsiva binaural)
CAD	Computer-aided design (Diseño asistido por computadora)
CINTRA	Centro de Investigación y Transferencia en Acústica
CPU	Central processing unit (Unidad central de procesamiento)
DMA	Direct memory access (Acceso directo a memoria)
DOF	Degrees of freedom (Grados de libertad)
DSP	Digital signal processor (Procesador digital de señales)
DXF	Drawing exchange format (Formato de intercambio de dibujo)
EDT	Early decay time
FDN	Feedback delay network (Red de retardo realimentada)
FFT	Fast Fourier transform (Transformada rápida de Fourier)
FIFO	First in, first out (Primero en entrar, primero en salir)

FIR	Finite impulse response (Respuesta impulsiva finita)
GPU	Graphics processing unit (Unidad de procesamiento gráfico)
HAL	Hardware abstraction layer (Capa de abstracción de hardware)
HRIR	Head-related impulse response (Respuesta impulsiva de cabeza)
HRT	Hard real-time
HRTF	Head-related transfer function (Función de transferencia de cabeza)
IIR	Infinite impulse response (Respuesta impulsiva infinita)
ILD	Interaural level difference (Diferencia de nivel interaural)
ISM	Image-source method (Método de fuente-imagen)
ISO	International organization for standardization (Organización internacional de normalización)
ITD	Interaural time difference (Diferencia de tiempo interaural)
jnd	Just-noticeable difference (Diferencia apenas notable)
LED	Light-emitting diode (Diodo emisor de luz)
LEV	Envolvimiento del oyente (Listener envelopment)
LTI	Linear time-invariant (Lineal e invariante en el tiempo)
LXRT	Linux real-time
MAA	Minimum audible angle (Mínimo ángulo audible)
MAMA	Minimum audible movement angle (Mínimo ángulo audible de una fuente en movimiento)
PC	Principal component (Componente principal)
PCA	Principal component analysis (Análisis de componentes principales)
PCI	Peripheral component interconnect (Interconexión de componentes periféricos)
POSIT	Pose from orthography and scaling with iterations
RIR	Room impulse response (Respuesta impulsiva del recinto)
RMS	Root mean square (Valor cuadrático medio)
RTAI	Real-time application interface (Interfaz para aplicaciones en tiempo real)

RTOS	Real-time operating system (Sistema operativo de tiempo real)
SD	Spectral distortion (Distorsión espectral)
SIMD	Single instruction, multiple data (Una instrucción, múltiples datos)
SMI	System management interrupt (Administración de señales de interrupción)
SNR	Signal-noise ratio (Relación señal-ruido)
SPL	Sound pressure level (Nivel de presión sonora)
SRT	Soft real-time
TSL	Total system latency (Latencia total del sistema)
USB	Universal serial bus (Bus universal en serie)
VS	Virtual source (Fuente virtual)

1 – Introducción

La realidad virtual (virtual reality, VR) tiene como objetivo crear situaciones en las cuales los humanos experimentan sensaciones que no se corresponden con su entorno físico actual, pero sí con uno virtual. Para arribar a esta meta, modelos computacionales generan estímulos compatibles con cada modalidad sensorial y estimulan a los sujetos mediante los transductores apropiados, como pantallas y auriculares. Actualmente, la VR tiene en cuenta los sentidos de visión, audición y tacto.

Históricamente, los principales esfuerzos en investigación y desarrollo en VR se han orientado hacia la computación gráfica. Con frecuencia, la simulación acústica era sólo un agregado, que tenía por finalidad producir efectos sonoros, sin correspondencia con el escenario simulado (Lentz 2007).

Sin embargo, en los últimos años, la *realidad acústica virtual* (acoustic virtual reality, AVR), también conocida como *simulación acústica virtual en tiempo real*, ha despertado interés en diversas áreas de la ciencia y la tecnología. Kleiner et al. (1993) señalaron que “la vivencia auditiva que un oyente experimentaría en un espacio dado, real o imaginario, podrá ser evocada si se sintetizan (mediante modelos físico-matemáticos) y se reproducen en los tímpanos del oyente las ondas sonoras que se originarían en el espacio modelado”. Es decir, 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*, o síntesis binaural, es la base que sustenta a la AVR y hace referencia a la representación auditiva del espacio. Es análogo al término “visualización” utilizado para la simulación de imágenes a partir de modelos en tres dimensiones.

La creación de un entorno auditivo virtual (auditory virtual environment, AVE), esto es, la componente auditiva de un entorno virtual, involucra tanto a aspectos físicos como a aspectos perceptuales. Cuando se habla de *evento sonoro* se hace referencia a los aspectos físicos del fenómeno de la audición, mientras que cuando se habla de *evento auditivo* se hace referencia a los aspectos subjetivos involucrados (Novo 2005). No puede especificarse una única relación causal entre un evento sonoro y un evento auditivo. Dos oyentes no siempre experimentarán (o escucharán) lo mismo cuando estén expuestos a la misma señal sonora. Aspectos perceptuales y cognitivos pueden influenciar lo que el sujeto escucha (Blauert 1997). Teniendo esto en cuenta, es necesario asumir que, bajo condiciones específicas, un evento sonoro evoca un único evento auditivo.

Un AVE puede ser materializado bajo diferentes abordajes pero, en general, se pueden distinguir tres. Uno de ellos es el llamado abordaje *físico-acústico* (o *auténtico*) cuyo propósito es conseguir una reproducción auténtica de un entorno, ver por ejemplo, Kuttruff (1995) y Botteldooren (1995). En este caso, el objetivo es evocar en el oyente la misma percepción que experimenta en el ambiente real. Apela a modelar de una manera exacta la propagación del sonido en el aire y sus reflexiones en los objetos del entorno. Debido a su complejidad, demanda grandes tiempos de cálculo.

Otro abordaje es el *perceptual* que apunta a manipular aquellos parámetros de las ondas sonoras que estimulan la percepción de ciertas características del entorno, ver por ejemplo, Blauert et al. (2000), Djelani et al. (2000) y Pellegrini (2002). Cuando una persona se expone a un campo sonoro real en un espacio cerrado, espontáneamente y de inmediato adquiere una imagen conceptual (conocida como impresión espacial) acerca del tipo, tamaño y propiedades del mismo.

Por último, el tercer abordaje es el *físico-perceptual* (o *plausible*) que tiene por propósito, evocar los mismos eventos auditivos que en un entorno real, ver por ejemplo, Savioja et al. (1999) y Wenzel et al. (2000). Se procura simplificar el abordaje físico-acústico apelando a evidencias perceptuales, teniendo en cuenta aquellas características que son relevantes y necesarias para una aplicación en particular, según los objetivos que se persiguen.

Los primeros sistemas buscaban una representación auténtica y totalmente fiel de la realidad pero, desafortunadamente, requerían un alto costo computacional y no fueron implementados con éxito. En la actualidad, las investigaciones se han inclinado hacia el desarrollo de entornos plausibles, donde se persigue lograr una percepción semejante en lugar de intentar reproducir la totalidad de las características físicas-acústicas de un ambiente real determinado. Pellegrini (2002) sugiere que no se puede alcanzar una auralización auténtica y que su calidad debe ser evaluada desde la perspectiva de su aplicación en particular (plausibilidad).

La expectativa que un usuario tiene de un sistema de simulación acústica es la de poder sentir y actuar como lo haría en un entorno natural. Por ejemplo, cuando una persona está inmersa en un campo sonoro tiende a girar su cabeza. Esta actitud inconsciente y espontánea ayuda a identificar la dirección de un sonido. Pequeños movimientos son suficientes para resolver ambigüedades de la posición de

una fuente sonora (Perrott y Saberi 1990; Arias et al. 2006; Arias et al. 2007). En consecuencia, para producir una convincente ilusión en tres dimensiones y evocar una imagen auténtica del espacio modelado, el sistema debe ser dinámico e interactivo. Dicho de otro modo, debe ser capaz de responder a los cambios del escenario acústico que se producen por una acción del usuario, en un tiempo perceptualmente aceptable y con una velocidad de actualización que garantice cierta continuidad.

Dos requisitos son importantes en estos sistemas de simulación interactivos: 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 está por debajo de un umbral y la latencia es excesiva, el oyente percibirá una imagen auditiva discontinua y retrasada, respectivamente.

En la década del 90, el uso de procesadores digitales de señales (digital signal processors, DSPs) ha permitido realizar un rápido procesamiento del audio. Éstos requieren una programación específica, por lo tanto, la portabilidad a otros sistemas implica un gran esfuerzo. En los últimos años, con el advenimiento de la computación moderna, el incremento en la capacidad de cálculo de las computadoras personales abrió la posibilidad para comenzar a implementar la AVR en plataformas estándares. Esto tiene la ventaja de bajar los costos de hardware y lograr una evidente mejoría en la portabilidad. La evolución mencionada, más los avances en la investigación sobre percepción multimodal, hará que en un futuro las fronteras entre una situación real y su correlato virtual sean prácticamente imperceptibles.

1.1 Subsistemas a modelar

En todo sistema de simulación acústica se deben modelar tres subsistemas: a) la fuente sonora; b) el recinto y el medio de propagación; y c) el oyente (Figura 1.1).

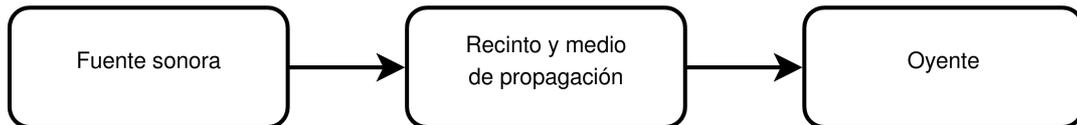


Figura 1.1 Diagrama de los tres subsistemas de un sistema de realidad acústica virtual

Las ondas de presión generadas por una fuente sonora viajan a través del medio y sufren un sinnúmero de transformaciones antes de alcanzar los tímpanos de un oyente. En primer lugar, las superficies y objetos del entorno provocan reflexiones y obstrucciones de las ondas alterando su recorrido y propiedades. En segundo lugar, son modificadas por el cuerpo del oyente.

La distribución de presión en la cercanía de una fuente sonora (sea natural o artificial) varía según la frecuencia y el espacio. Esta distribución se caracteriza por el diagrama de radiación, conocido como directividad de la fuente.

La transmisión del sonido en un recinto, entre una fuente y un oyente, está influenciada por una serie de factores: el medio de propagación (aire), las dimensiones y las características físico-acústicas de las superficies y objetos, y la posición de la fuente y el oyente. Esta transmisión se caracteriza, en el dominio del tiempo, por la respuesta impulsiva del recinto (room impulse response, RIR).

Las ondas sonoras que alcanzan los tímpanos de un oyente son transformadas por la cabeza, los pabellones auriculares (oído externo), los hombros y el torso. Estas transformaciones tanto temporales como espectrales son diferentes para cada

oído y dependen del ángulo de incidencia de la onda. Las respuestas impulsivas de cabeza (head-related impulse responses, HRIRs), en el dominio del tiempo, o las funciones de transferencia de cabeza (head-related transfer functions, HRTFs), en el dominio de la frecuencia, contienen y caracterizan las transformaciones mencionadas.

La directividad de la fuente, junto a la respuesta impulsiva del recinto y a las funciones de transferencia del oyente forman la respuesta impulsiva binaural (binaural impulse response, BIR) (una para cada oído), que caracteriza la transmisión global entre una fuente y los tímpanos de dicho oyente. De esta manera, la sensación de presencia en el entorno modelado podrá ser evocada en el oyente si mediante auriculares se reproduce la señal que resulta de la convolución de la BIR con una señal anecoica (libre de reflexiones).

Los subsistemas mencionados constituyen, en sí mismos, tópicos de estudio bien diferenciados en acústica y sus resultados contribuyen a enriquecer los AVEs, los cuales estarán limitados por el grado de avance de la investigación en audición espacial.

1.2 Aplicaciones de los sistemas de simulación acústica

Los campos de investigación relacionados con estos sistemas son multidisciplinarios y requieren conocimientos sobre física acústica, acústica de recintos, procesamiento digital de señales, sistemas de tiempo real y psicoacústica.

Hay aplicaciones que requieren estimular diferentes aspectos de la experiencia auditiva del usuario y que demandan la ejecución de tareas diversas. Por ejemplo, si un diseñador de auditorios quisiera simular para su cliente cómo escucharía una sala diseñada antes de que sea construida, será preciso que el sistema realice una

síntesis exacta de las características de la misma. En una aplicación de interfaz hombre-máquina, en cambio, será más importante localizar con precisión la ubicación de una fuente sonora que la sensación de inmersión. En el caso de entrenadores de vuelo o combate, será crucial simular con precisión tanto la información direccional como la sensación de presencia en el entorno modelado.

Los sistemas que simulan AVEs tradicionalmente se han aplicado a: el diseño y evaluación de recintos, principalmente en salas de conciertos; la investigación psicoacústica, ya que se pueden controlar a voluntad los parámetros de los estímulos y estudiar su implicancia perceptual; la reproducción multicanal de audio, como los sistemas Dolby; los simuladores de vuelo y combate; las aplicaciones multimedia en animaciones y juegos; teleconferencias; y en tecnología de interfaz hombre-computadora.

El mayor valor de los AVEs es permitir la investigación del comportamiento humano en contextos realistas e interactivos controlados, sin las limitaciones de orden físico y económico para construirlos. Además, la VR, en general, y la AVR, en particular, son herramientas invaluable para el entrenamiento de habilidades en ambientes seguros y libres de peligros.

1.3 Objetivo y alcance de la tesis

La tendencia en el desarrollo de los sistemas de simulación acústica virtual estuvo siempre limitada a su aplicación, carente de flexibilidad y posibilidades de expansión. Día a día, sus posibilidades de aplicación se diversifican, alcanzando a áreas de la ciencia y la tecnología impensadas hace algunos años atrás. Estos hechos plantean un dilema al momento de encarar el desarrollo de un sistema que simula un AVE.

La finalidad del sistema de simulación acústica virtual en tiempo real que se presenta en esta tesis, denominado *AVRS*, es constituir una genuina plataforma experimental flexible que, por un lado, facilite la investigación de aspectos aún no resueltos de la simulación acústica y, por el otro, permita su utilización como un verdadero banco de pruebas en nuevas aplicaciones no tradicionales. También es de interés que el AVRS se ejecute sobre una arquitectura de computadora de propósito general (con hardware y software estándares).

El objetivo específico de la tesis es:

Desarrollar un sistema experimental dinámico e interactivo de realidad acústica virtual, basado en arquitectura de computadoras de propósito general y que sea flexible, escalable, fácil de mantener y modular.

Para alcanzar este objetivo se sigue el abordaje físico-perceptual, que permite lograr modelos simplificados factibles de operar en tiempo real, basados en evidencias perceptuales. Es decir, se busca sintetizar sonidos mediante métodos físico-matemáticos que, al ser escuchados mediante auriculares, produzcan en el oyente la sensación de inmersión en el espacio simulado.

1.4 Organización del informe de tesis

La tesis ha sido dividida en ocho capítulos. En el presente, se introdujo al lector en los principales conceptos relativos a la realidad acústica virtual, los antecedentes, las aplicaciones y se presentó el objetivo y alcance de la tesis.

En el Capítulo 2 se exponen consideraciones teóricas respecto a los aspectos acústicos y psicoacústicos involucrados en los modelos utilizados a lo largo de la

tesis.

El Capítulo 3 presenta el modelo de la fuente sonora y la forma de implementación de sus patrones de directividad.

En el Capítulo 4, se discute el modelo del recinto y el medio de transmisión. Específicamente, se detalla la simulación de la RIR.

El Capítulo 5 trata sobre el modelo del oyente, es decir, cómo se transforma una señal que representa a la fuente y al recinto en otra binaural que llega a los tímpanos del oyente. También se realiza una breve discusión acerca de diferentes métodos de personalización de HRTFs a partir de mediciones de HRTFs genéricas.

En el Capítulo 6 se integran todos los modelos descritos mediante el desarrollo de un sistema interactivo basado en software de simulación acústica virtual en tiempo real.

En el Capítulo 7 se presentan los resultados de un caso de estudio como validación del sistema.

Por último, en el Capítulo 8, se resumen los resultados obtenidos y se enuncian las conclusiones del trabajo de tesis.

Además, acompañan a este informe, dos apéndices en los que se incluye información adicional que puede ser de utilidad para aclarar algunos tópicos.

En el Apéndice A se definen los parámetros acústicos utilizados que caracterizan al campo sonoro de un recinto, los cuales están especificados en la norma ISO 3382 (2009).

El Apéndice B está dedicado a los sistemas de coordenadas de la cabeza más comunes y que se utilizan en diferentes partes de la tesis.

2 – Fundamentos de acústica y psicoacústica

El estudio de la física acústica resulta esencial para comprender las leyes que rigen la propagación del sonido y modelar la transmisión de las ondas acústicas en ambientes abiertos, cerrados o semi cerrados. Por otra parte, la principal función de la audición es determinar las características de la fuente sonora a partir de la información contenida en los sonidos que ella produce (Yost 1991). El sistema auditivo le permite al ser humano realizar minuciosos juicios sobre localización, calidad y confort acústico, entre otros. Comprender los diferentes aspectos acústicos y psicoacústicos involucrados en la audición resulta clave para modelar y programar la auralización, determinar su precisión y establecer sus aspectos fundamentales.

2.1 Propagación de las ondas sonoras

Una propiedad importante de las ondas sonoras es que transportan energía e información. Cuando una onda se propaga, se producen cambios en la presión y en la densidad del medio. Luego de que la onda lo atraviesa, el medio retorna a su

estado inicial. Todo campo sonoro complejo puede ser considerado, en principio, como la superposición de numerosas ondas sonoras elementales. A continuación se explican conceptos relacionados a la propagación de estas ondas sonoras.

2.1.1 Velocidad del sonido

Si la propagación es libre de pérdidas, el medio es homogéneo y sin límites en todas direcciones, la velocidad de propagación del sonido es constante con respecto al espacio y al tiempo (Kuttruff 2009). En general, la velocidad de propagación c , es función de la densidad ρ y del coeficiente de elasticidad E , siendo

$$c = \sqrt{\frac{E}{\rho}} \quad \left[\frac{\text{m}}{\text{s}} \right], \quad (2.1)$$

donde ρ está dado en kg/m^3 y E en N/m^2 .

Ahora, la velocidad de propagación del sonido en el aire depende de la temperatura y se puede calcular a partir de la siguiente expresión

$$c = 331,4 \sqrt{1 + \frac{\Theta}{273}} \quad \left[\frac{\text{m}}{\text{s}} \right], \quad (2.2)$$

donde Θ es la temperatura en grados centígrados. Si se consideran condiciones normales y 20°C de temperatura, la velocidad de propagación del sonido en el aire es de aproximadamente 343 m/s .

En un rango de aproximadamente $\pm 30^\circ\text{C}$, dicha fórmula se puede aproximar mediante

$$c = 331,4 + 0,6\Theta \quad \left[\frac{\text{m}}{\text{s}} \right]. \quad (2.3)$$

En los recintos cerrados, debido a la variación de temperatura y a la influencia del aire acondicionado, no se puede considerar que el aire está en reposo¹, por lo

1 Estrictamente hablando, el aire nunca se encuentra en reposo (equilibrio estático). Aunque no existan corrientes de aire, las moléculas se mueven aleatoriamente por efecto de la temperatura.

que esta ecuación no se cumpliría. Kuttruff (2009) indica que los efectos causados por estos cambios son pequeños y pueden ser descuidados.

2.1.2 Ecuación de onda

Una onda sonora puede ser descripta completamente mediante los desplazamientos instantáneos de las partículas. Sin embargo, a menudo, se utiliza la velocidad de la partícula como una magnitud acústica básica, en lugar del desplazamiento. Las vibraciones en la onda sonora no suceden en todos los puntos con la misma fase. En ciertos puntos del campo sonoro las partículas vibran con fase opuesta. Por lo tanto, en algunas regiones las partículas se juntan y en otras tienden a separarse (Kuttruff 2009).

En un gas, bajo la influencia de una onda sonora, ocurren variaciones en su presión y densidad. Estas son funciones del tiempo y de la posición. La diferencia entre la presión instantánea y la presión estática se denomina *presión sonora*.

La onda sonora provoca variaciones de temperatura adiabáticas, ya que los cambios en la presión del gas ocurren generalmente a una velocidad tal que no existe intercambio de calor entre unidades de volumen adyacentes. Como consecuencia, la temperatura puede considerarse como un parámetro característico de la onda sonora.

Existen leyes básicas que conectan varios parámetros acústicos. Utilizando estas leyes es posible plantear una ecuación diferencial general que rige la propagación del sonido en una, dos y tres dimensiones (Beranek 1986).

En primer lugar, debe definirse la conservación del momento. Puede expresarse de la siguiente manera

$$\text{grad}(p) = -\rho_0 \frac{\partial \mathbf{v}}{\partial t} \quad (2.4)$$

donde grad denota al gradiente, p es la presión sonora, ρ_0 es el valor estático de la densidad del gas, \mathbf{v} es la velocidad de la partícula, y t es el tiempo.

La conservación de la masa conduce a la siguiente expresión

$$\rho_0 \text{div}(\mathbf{v}) = -\frac{\partial \rho}{\partial t} \quad (2.5)$$

donde div denota a la divergencia, ρ es la densidad total del gas, es decir, $\rho = \rho_0 + \delta\rho$. Se asume que los cambios de p y ρ son pequeños comparados con los valores estáticos p_0 y ρ_0 . También se asume que la velocidad de la partícula \mathbf{v} es mucho menor a la velocidad del sonido.

Para un gas ideal, se mantienen las siguientes relaciones entre la presión sonora, las variaciones de la densidad del gas y los cambios de temperatura

$$\frac{p}{p_0} = \kappa \frac{\delta\rho}{\rho_0} = \frac{\kappa}{\kappa - 1} \frac{\delta\Theta}{\Theta + 273} \quad (2.6)$$

donde κ es un exponente adiabático. Despreciando \mathbf{v} y $\delta\rho$ en las ecuaciones (2.4) y (2.6) se puede obtener la siguiente ecuación diferencial

$$c^2 \Delta p = \frac{\partial^2 p}{\partial t^2} \quad (2.7)$$

con

$$c^2 = \kappa \frac{p_0}{\rho_0}. \quad (2.8)$$

Esta ecuación diferencial se denomina *ecuación de onda*, y rige la propagación de las ondas sonoras en cualquier fluido sin pérdidas (Kuttruff 2009).

2.1.3 Onda plana

La solución directa para la propagación del sonido en una dimensión es la llamada onda plana. Considerando sólo la dirección x de un sistema de coordenadas cartesiano, la ecuación (2.7) se puede expresar como

$$c^2 \frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial t^2}. \quad (2.9)$$

Esta ecuación diferencial de segundo orden tiene la siguiente solución general

$$p(x, t) = F(ct - x) + G(ct + x) \quad (2.10)$$

donde F representa una onda de presión desplazándose en la dirección positiva de x con velocidad c , y G describe una onda de presión sonora propagándose en la dirección negativa de x con la misma velocidad.

Cada término en esta ecuación representa una onda plana progresiva, donde la presión sonora p es constante en todo plano perpendicular al eje x . Estos planos de presión constante se denominan *frentes de onda* (Figura 2.1) (Mankovsky 1971; Kinsler et al. 1999).

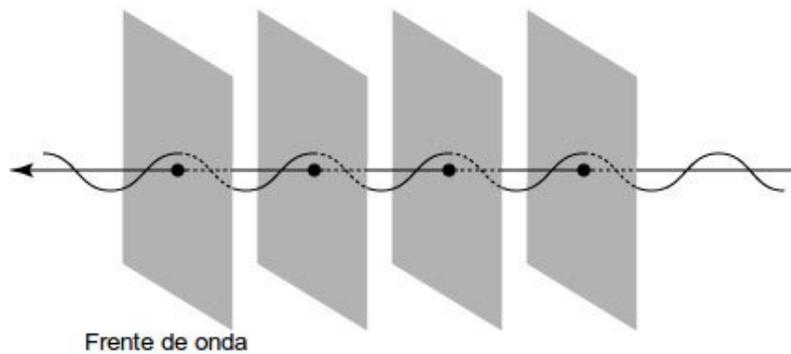


Figura 2.1 Frentes de onda de una onda plana

La velocidad de la partícula, como lo muestra la ecuación (2.4), tiene una sola componente no nula (la componente paralela al gradiente de presión sonora). Como consecuencia, las ondas sonoras en los fluidos son longitudinales. A partir de las ecuaciones (2.4) y (2.10) se puede obtener la siguiente expresión de la velocidad de la partícula

$$v = v_x = \frac{1}{\rho_0 c} [F(ct - x) - G(ct + x)]. \quad (2.11)$$

La razón entre la presión sonora y la velocidad de la partícula en la onda plana que se propaga en dirección positiva ($G = 0$), es independiente de la frecuencia como se puede observar en las ecuaciones (2.10) y (2.11), entonces

$$\frac{p}{v} = \rho_0 c = Z_0. \quad (2.12)$$

El valor Z_0 se denomina *impedancia característica de la onda plana* o *impedancia de la onda*. Se puede interpretar como la cantidad de presión necesaria para lograr que las partículas del medio se pongan en movimiento.

2.1.4 Onda esférica

En las ondas esféricas, los frentes de onda (superficies de presión constante) son esferas concéntricas (Figura 2.2). En el centro de estas esferas se puede imaginar una fuente sonora puntual, infinitesimalmente pequeña. Debido a esto, conviene utilizar coordenadas polares.

Transformando la ecuación diferencial (2.7) a coordenadas polares, queda de la siguiente manera

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r} \frac{\partial p}{\partial r} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}. \quad (2.13)$$

Una solución a esta ecuación, propuesta por Kuttruf (2009), es

$$p(r, t) = \frac{\rho_0}{4\pi r} Q' \left(t - \frac{r}{c} \right). \quad (2.14)$$

donde Q es el volumen de la velocidad o “flujo de sonido”.

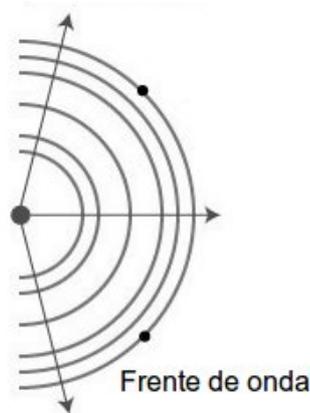


Figura 2.2 Frentes de onda de una onda esférica (vista en dos dimensiones)

Esta es la expresión general para una onda esférica producida por una fuente puntual en r con flujo de sonido $Q(t)$, siendo Q' su derivada con respecto al tiempo. El argumento $t - r/c$ indica que cualquier perturbación creada por la fuente sonora se propaga con velocidad c y su amplitud decrece en un factor $1/r$.

2.1.5 Energía y densidad de energía

Cuando una fuente sonora genera una onda, entrega energía que se propaga con dicha onda. La energía de una señal sonora limitada en el tiempo se expresa de la siguiente manera

$$E = \int_0^T p^2(t) dt \quad (2.15)$$

donde T es el periodo de observación de la señal, p es la presión y t es el tiempo (Beranek 1986).

La energía contenida en una unidad de volumen está caracterizada por una magnitud llamada *densidad de energía*. Se debe distinguir entre densidad de

energía potencial y densidad de energía cinética, que se expresan de la siguiente manera

$$w_{pot} = \frac{p^2}{2\rho_0 c^2} \quad (2.16)$$

$$w_{cin} = \frac{\rho_0 |\mathbf{v}|^2}{2}. \quad (2.17)$$

Entonces, la densidad de energía total es

$$w = w_{pot} + w_{cin}. \quad (2.18)$$

2.1.6 Nivel de presión sonora

Generalmente la presión sonora no se expresa en forma lineal, es decir, en pascales ($1 \text{ Pa} = 1 \text{ N/m}^2$), sino que se indica en una escala logarítmica donde la unidad de medida es el decibel (dB).

Una razón por lo que se hace esto es aproximarse a la respuesta del oído humano (Vorländer 2007). La otra razón es para hacer manejable el enorme rango de presiones sonoras existentes. Por ejemplo, asumiendo una presión estática de 100 kPa en la atmósfera al nivel del mar, las presiones sonoras irían de 0,00001 Pa a 1000 Pa. El límite inferior es cercano al “umbral de audición”, el límite de sensibilidad del oído humano, y el límite superior al “umbral del dolor”. En la práctica, este rango enorme se convierte a una escala logarítmica entre 0 dB y 130 dB aproximadamente.

El nivel de presión sonora (sound pressure level, SPL) o nivel sonoro es la medida logarítmica de la presión sonora efectiva de un sonido con respecto a un valor de referencia. El SPL se define por la siguiente expresión

$$L = 10 \log_{10} \frac{\tilde{p}^2}{p_0^2} = 20 \log_{10} \frac{\tilde{p}}{p_0} \quad [\text{dB}] \quad (2.19)$$

donde \tilde{p} es la presión sonora efectiva o RMS (root mean square), definida como

$$\tilde{p} = \sqrt{\frac{1}{T} \int_0^T p^2(t) dt} \quad (2.20)$$

y la referencia de presión sonora efectiva en el aire p_0 es de $20 \mu\text{Pa}$ ($20 \times 10^{-6} \text{Pa}$), la cual es considerada el umbral del oído humano (a 1 kHz) (Kuttruff 2009).

2.1.7 Intensidad sonora

Otra magnitud importante es la intensidad sonora. Es una medida de la energía transportada por una onda sonora. Se define como la cantidad de energía por segundo que pasa a través de una sección plana de 1 m^2 perpendicular a la dirección de propagación.

El vector de intensidad sonora se expresa como

$$\mathbf{I} = p\mathbf{v} \quad (2.21)$$

donde p es la presión sonora y \mathbf{v} es la velocidad de la partícula.

En una onda plana, la presión sonora y la velocidad de la partícula están en fase, ecuación (2.12), y la fórmula de intensidad sonora se reduce a

$$|\mathbf{I}| = \frac{\tilde{p}^2}{\rho_0 c} \quad (2.22)$$

con la dirección de la intensidad sonora en la dirección de propagación.

La intensidad sonora también puede expresarse con un nivel, el nivel de intensidad

$$L_I = 10 \log_{10} \frac{|\mathbf{I}|}{I_0} \quad (2.23)$$

donde $I_0 = 10^{-12} \text{ W/m}^2$.

2.2 Leyes de la reflexión sonora

En la realidad todas las ondas sonoras originadas en una fuente acústica son

esféricas. La reflexión de una onda esférica en una pared no rígida es compleja. Sin embargo, si la fuente sonora está lo suficientemente alejada de la superficie de reflexión, el fenómeno se puede simplificar despreciando la curvatura del frente onda. Se puede asumir que la onda incidente es plana y sin perturbaciones.

Entonces, se considera el caso en que la onda incidente es plana y que la superficie de reflexión es uniforme y se extiende infinitamente. Generalmente, en tal situación, parte de la energía sonora incidente es reflejada como si una onda plana fuese generada en la superficie de incidencia. Pero la onda reflejada difiere en amplitud y fase con la onda incidente. La onda reflejada y la incidente se interfieren mutuamente formando (al menos parcialmente) una onda estacionaria².

2.2.1 Factor de reflexión y coeficiente de absorción

Los cambios en amplitud y fase pueden ser expresados por un factor de reflexión complejo R que es propio de cada superficie

$$R = |R|e^{i\chi} \quad (2.24)$$

donde $|R|$ es el valor absoluto y χ el ángulo de fase, los cuales dependen de la frecuencia y la dirección de incidencia de la onda.

Como la intensidad de la onda plana es proporcional al cuadrado de la presión, la intensidad de la onda reflejada es menor que la incidente en un factor de $|R|^2$. El coeficiente de absorción α de una superficie describe la cantidad de energía que se pierde durante una reflexión

$$\alpha = 1 - |R|^2 \quad (2.25)$$

2 Es una onda en la cual ciertos puntos, llamados nodos, permanecen inmóviles. Es producto de la interferencia de dos ondas.

2.2.2 Impedancia sonora de la superficie

Otro parámetro relacionado a las propiedades de la superficie de reflexión está basado en la velocidad de la partícula efectiva en dirección normal a la superficie y la presión sonora efectiva en la superficie. Este parámetro se denomina impedancia sonora de la superficie y está definida por

$$Z = \left(\frac{p}{v_n} \right)_{\text{superficie}} \quad (2.26)$$

donde p representa la presión sonora efectiva en la superficie y v_n la componente de velocidad de las partículas normal a la superficie.

La impedancia, al igual que el factor de reflexión, es una magnitud compleja, y es función del ángulo de incidencia y varía con la frecuencia.

2.2.3 Reflexión del sonido con incidencia normal

Assumiendo que la superficie se encuentra ubicada en posición normal a la dirección de propagación de la onda incidente y la onda se dirige a la superficie (donde $x = 0$) desde una dirección negativa de x (Figura 2.3), la presión sonora $p_i(x, t)$ y la velocidad de la partícula $v_i(x, t)$ son³

$$p_i(x, t) = \hat{p}_0 e^{i(\omega t - kx)} \quad (2.27)$$

$$v_i(x, t) = \frac{\hat{p}_0}{\rho_0 c} e^{i(\omega t - kx)} \quad (2.28)$$

donde \hat{p}_0 es el valor absoluto de la presión sonora inicial de la onda incidente, ω es la frecuencia angular, k es el número de onda ($k = \omega/c$), t es el tiempo, x es la coordenada en el eje de las abscisas, ρ_0 es el valor estático de la densidad del gas y c es la velocidad del sonido.

3 Usualmente se utiliza la notación exponencial por ser más compacta, con $e^{i\omega t} = \cos \omega t + i \sin \omega t$.

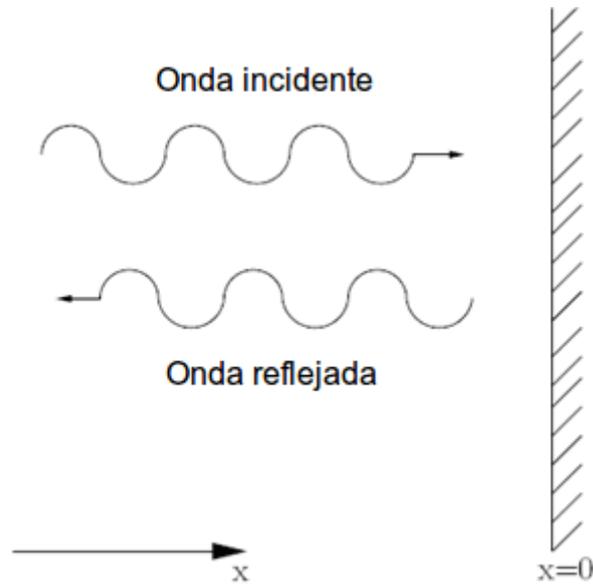


Figura 2.3 Reflexión del sonido con incidencia normal. Adaptado de Kuttruff (2009).

La onda reflejada tendrá menor amplitud y estará desfasada con respecto a la onda incidente. Ambos cambios son debido al factor de reflexión R . Además, el signo de la velocidad de la partícula cambia porque la onda reflejada se propaga en dirección contraria a la incidente. Entonces, la presión sonora $p_r(x, t)$ y la velocidad de la partícula $v_r(x, t)$ de la onda reflejada serán

$$p_r(x, t) = R\hat{p}_0 e^{i(\omega t + kx)} \quad (2.29)$$

$$v_r(x, t) = -R \frac{\hat{p}_0}{\rho_0 c} e^{i(\omega t + kx)}. \quad (2.30)$$

2.2.4 Reflexión del sonido con incidencia oblicua

Un caso más general es aquel en que el ángulo de incidencia es cualquiera. Ahora, se asume que la normal a la superficie y la dirección de incidencia de la onda son rectas contenidas en el plano xy representado en la Figura 2.4.

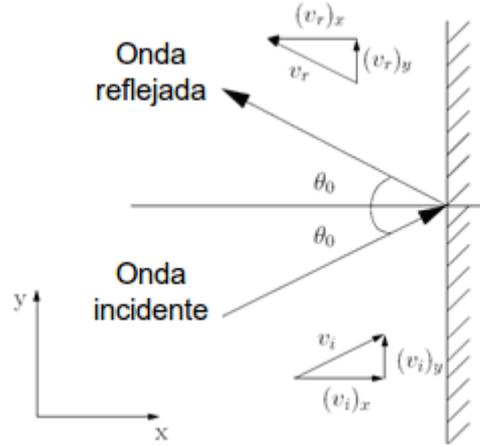


Figura 2.4: Reflexión del sonido con incidencia oblicua. Adaptado de Kuttruff (2009).

Redefiniendo x por x' de la siguiente manera

$$x' = x \cos \theta_0 + y \sen \theta_0 , \quad (2.31)$$

se crea un nuevo sistema de coordenadas donde los ejes están rotados un ángulo θ_0 con respecto al sistema xy . El resultado es una onda plana propagándose en la dirección positiva de x' .

Reemplazando en (2.27), queda

$$p_i = \hat{p}_0 e^{i(\omega t - kx \cos \theta_0 - ky \sen \theta_0)} . \quad (2.32)$$

La componente normal a la superficie de la velocidad (componente x) es necesaria para calcular la impedancia sonora de la superficie. Luego, se obtiene

$$(v_i)_x = \frac{\hat{p}_0}{\rho_0 c} \cos \theta_0 e^{i(\omega t - kx \cos \theta_0 - ky \sen \theta_0)} \quad (2.33)$$

Para este caso, donde la incidencia de la onda es oblicua, también se debe cambiar el signo de x en la expresión de la onda reflejada. Entonces, de forma similar a las ecuaciones (2.29) y (2.30), se obtiene la presión sonora y la velocidad de la partícula de la onda reflejada

$$p_r(x, t) = R\hat{p}_0 e^{i(\omega t + kx \cos \theta_0 - ky \sin \theta_0)} \quad (2.34)$$

$$(v_r)_x(x, t) = -R \frac{\hat{p}_0}{\rho_0 c} \cos \theta_0 e^{i(\omega t + kx \cos \theta_0 - ky \sin \theta_0)}. \quad (2.35)$$

2.3 Difracción, dispersión y reflexión difusa

Estas leyes de reflexión sonora sólo son válidas para superficies de extensión infinita. En un caso general, cualquier borde de una superficie dispersará algo de energía sonora en todas las direcciones. Este fenómeno es llamado difracción.

Kuttruff (2009) plantea que la difracción puede ser explicada observando la Figura 2.5. En la misma se ilustra un plano rígido con un borde recto. Si una onda plana incide en el plano con dirección normal, es de esperar que el plano refleje una parte de la onda a la región A. Al mismo tiempo no se emiten reflexiones hacia la región B, a la que llamaremos zona de sombra. Esto sería cierto si la longitud de onda del sonido incidente es infinitesimalmente pequeña. Sin embargo, en un caso real, una onda provocada por la difracción en el borde del plano estará presente en la zona de sombra. Por otra parte, en la región C, la onda plana se distorsiona debido a interferencias con la onda de difracción. Resumiendo, el límite entre la zona de sombra (B) y la zona iluminada (C) no es abrupto, sino difuso debido a la onda de difracción. Un efecto similar ocurre en el límite superior de la región A.

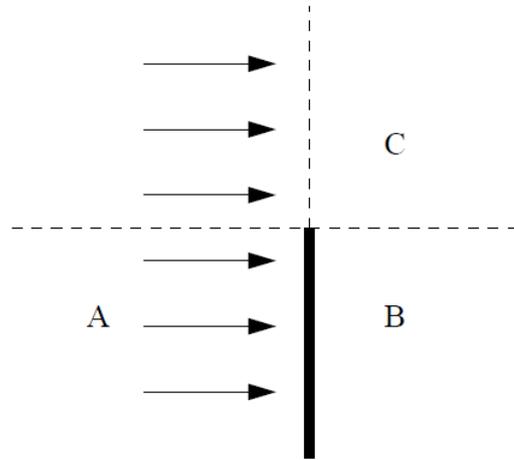


Figura 2.5 Difracción de una onda plana en el borde de un plano rígido. Extraído de Kuttruff (2009).

Generalmente, una superficie cualquiera, no es totalmente plana sino que contiene múltiples irregularidades. Si éstas son muy pequeñas con respecto a la longitud de onda del sonido incidente, la reflexión especular no se verá afectada. Por el contrario, si las irregularidades son grandes comparadas a la longitud de onda, cada una de estas caras irregulares pueden ser tratadas como secciones de superficies y las ondas incidentes se pueden reflejar especularmente. Sin embargo, existe un rango intermedio de longitudes de onda donde cada cara agrega dispersión a la reflexión especular de toda la superficie. Por lo tanto, si la superficie tiene una estructura irregular una gran parte de la energía sonora será dispersada en todas direcciones (Figura 2.6). El término reflexión difusa en la superficie también suele ser usado en este contexto.

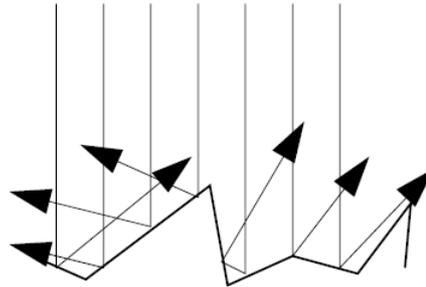


Figura 2.6 Reflexión difusa en una superficie de estructura irregular.

Extraído de Kuttruff (2009).

Cabe aclarar que en acústica de recintos el término difusión puede ser atribuido a dos conceptos diferentes. Por una parte la difusión es una propiedad del campo sonoro. En él se describe una isotropía de la uniformidad de dirección de la propagación del sonido (ver siguiente sección). Por otra parte, la difusión es la característica de una superficie para dispersar el sonido incidente en múltiples direcciones. Si bien el campo sonoro difuso puede ser consecuencia de la reflexión difusa en bordes, se deben diferenciar ambos fenómenos.

Como se puede apreciar, el término dispersión está relacionado con la difracción y también con la reflexión difusa. Dalenbäck et al. (1993) han definido estos términos de la siguiente manera:

- **Difracción:** Teóricamente y en una visión microscópica de la onda, la difracción es una de las causas de la reflexión difusa. Comúnmente se relaciona con el efecto producido en los bordes de un elemento reflector.
- **Dispersión:** Este término es generalmente utilizado en acústica lineal para describir el resultado de la difracción. Se relaciona, de una manera muy general, a la reflexión en una superficie áspera.
- **Reflexión difusa:** Es el término mas apropiado para describir el proceso de

reflexión en una superficie difusora.

2.4 Campo sonoro difuso

Un campo sonoro típico en un espacio cerrado está compuesto por la sumatoria de varias ondas planas. Cada una de ellas tiene un valor de amplitud, fase y una dirección determinada. Para conocer el efecto de la superficie de reflexión en tan complicado campo sonoro, se deben considerar las reflexiones de cada onda plana por separado. Luego se deben sumar todos los valores de presión sonora de estas ondas.

Otra posibilidad es valerse de algunas simplificaciones. En primer lugar se desprecian todas las relaciones de fases y sus efectos de interferencia. Esto es posible si se asume que las fases de las ondas incidentes están distribuidas aleatoriamente. En esta situación es suficiente sumar o promediar las energías de las ondas. Estos valores son proporcionales al cuadrado de las presiones de las ondas elementales. En segundo lugar si se asume que las intensidades del sonido incidente están uniformemente distribuidas en todas las direcciones posibles, cada porción de ángulo sólido lleva la misma cantidad de energía por segundo hacia la superficie reflejante. Esto se denomina incidencia sonora aleatoria. El campo sonoro asociado a la situación descrita se denomina *campo sonoro difuso* (Miles 1984; Kuttruff 1995; Kuttruff 2009).

2.5 Reverberación

En un recinto cerrado, cuando se interrumpe el sonido emitido por una fuente sonora, existe un tiempo finito para que la energía sonora decaiga a valores inaudibles. Este remanente de sonido luego de que la señal de excitación fue interrumpida se llama *reverberación*.

Con el fin de poder cuantificar la reverberación de un recinto, se define el *tiempo de reverberación* (para una frecuencia determinada), como el tiempo que transcurre desde que la emisión de una fuente sonora se detiene hasta el momento en que la energía disminuye 60 dB con respecto a su valor inicial. Esto representa un cambio en la intensidad sonora de 1000000 de veces ($10 \log_{10} 10^6 = 60$ dB), o un cambio en la presión sonora de 1000 veces ($20 \log_{10} 10^3 = 60$ dB).

Este parámetro fue introducido por W. C. Sabine a finales del siglo XIX quien, a partir de algunos experimentos, estimó la siguiente relación entre el tiempo de reverberación del recinto, su volumen y su absorción total

$$T_{60} = 0,161 \frac{V}{S\bar{\alpha}} \quad (2.36)$$

donde T_{60} es el tiempo de reverberación en segundos, V el volumen del recinto en m^3 , S el área total del recinto en m^2 , y $\bar{\alpha}$ el coeficiente de absorción promedio en sabins. Este último valor es el promedio ponderado por superficie de los coeficientes de absorción individuales (Kinsler et al. 1999)

$$\bar{\alpha} = \frac{1}{S} \sum_i S_i \alpha_i \quad (2.37)$$

donde α_i es el coeficiente de absorción de la i -ésima superficie de área S_i .

Como se dijo, no sólo las superficies del recinto absorben la energía del sonido, sino también el aire. En el caso de que se trate de recintos muy grandes, como auditorios o iglesias, el tiempo que las ondas sonoras transcurren en el aire es mayor. La ecuación (2.36) se debe modificar para tener en cuenta esta absorción y lograr resultados más precisos (Beranek 1986). Entonces, queda

$$T_{60} = 0,161 \frac{V}{S\bar{\alpha} + 4mV} \quad (2.38)$$

donde m es una constante de atenuación de energía en metros^{-1} .

Estas ecuaciones son útiles cuando la absorción total de la sala no es muy grande. Norris y Eyring (Eyring 1930) propusieron otra expresión para el cálculo del tiempo de reverberación que brinda resultados más próximos a lo que ocurre en la realidad. La fórmula de Norris-Eyring es

$$T_{60} = -0,161 \frac{V}{S \ln(1 - \bar{\alpha})} . \quad (2.39)$$

Por lo general, el tiempo de reverberación tiende a disminuir a medida que la frecuencia aumenta. Esto es debido a las características de mayor absorción que tienen los materiales comúnmente empleados en las altas frecuencias y también a la absorción del aire, especialmente en recintos grandes.

2.6 Psicoacústica: audición espacial

El sistema auditivo es un sofisticado procesador espacial, que le permite al organismo detectar y discriminar la ubicación de objetos sonoros en el espacio (Arias y Ramos 2003). Grantham (1995) señala que un oyente tiene una inmediata apreciación del espacio auditivo⁴, en el sentido de que su orientación hacia el evento sonoro es natural, rápida y, en general, exacta.

Aunque la precisión espacial es más pobre en el dominio auditivo que en el visual (de hasta dos órdenes de magnitud), el “mundo auditivo” tiene la ventaja de extenderse en todas las direcciones alrededor del observador, mientras que el “mundo visual” se restringe sólo a su región frontal (siendo óptima la resolución visual dentro de unos pocos grados de la línea de mirada). En situaciones de alta incertidumbre espacial, esto es, cuando un observador no sabe para dónde mirar, la potencia del sistema auditivo para proveer información espacial es excelente (Arias

4 Espacio que se extiende alrededor de la cabeza en todas las direcciones en presencia de un evento auditivo.

y Ramos 2003). Perrott et al. (1990) han sugerido que esta geometría perceptual diferencial permite que el sistema auditivo reaccione primero a los eventos ambientales y “dirija” la cabeza del observador para que apunte hacia ellos, de manera tal que entre en juego el sistema visual para realizar un análisis espacial más refinado.

Las grandes temáticas de este área son: el procesamiento binaural, la localización sonora (sonidos directos) y el efecto precedente (localización de sonidos reflejados).

2.6.1 Localización de fuentes sonoras

La localización sonora se refiere a la percepción de la posición de una fuente sonora en el plano horizontal o azimutal (izquierda-derecha) y en el plano medio o vertical (arriba-abajo), y a la percepción de la distancia relativa entre el oyente y la fuente (Blauert 1997).

Una gran parte de los estudios de localización auditiva se han realizado en condiciones artificiales, por ejemplo en ambientes anecoicos, sin que el participante pudiera moverse libremente y con estímulos artificiales como tonos puros, clics o *bursts* de ruido. Si bien estas condiciones han permitido conocer la mayor parte de lo que se sabe sobre audición espacial, explican muy poco sobre cómo una persona funciona auditivamente en la vida real y cotidiana (Brown y May 2005).

2.6.2 Claves involucradas en la localización

La experimentación realizada tanto en campo libre como con auriculares ha permitido identificar las principales claves que intervienen cuando una persona localiza sonidos (Grantham 1995; Blauert 1997).

Las claves binaurales son: a) la diferencia de tiempo interaural (interaural time

difference, ITD), y b) la diferencia de nivel interaural (interaural level difference, ILD).

La ITD es la diferencia de tiempo con que un sonido llega a cada uno de los oídos. Para fuentes sonoras ubicadas fuera del plano medio (azimut diferente a 0°), el sonido alcanza primero al oído más cercano a la fuente. En adultos, la ITD varía entre 0 y $700 \mu\text{s}$ para fuentes ubicadas a 0° y 90° de azimut respectivamente (Basso 2006).

Cuando la fuente sonora se ubica fuera del plano medio, la intensidad de los sonidos que llegan a cada uno de los oídos también es ligeramente diferente. La ILD se refiere a estas variaciones de intensidad ocasionadas por la cabeza y el torso, que actúan como una barrera acústica que se interpone entre la fuente y el oído más alejado de ésta.

En el plano horizontal, la posición percibida de la fuente sonora depende tanto del valor de ITD como de ILD.

Por otro lado, el efecto de filtrado natural es una clave monoaural que ayuda en la localización de una fuente sonora. La cabeza y los pabellones auditivos, de forma irregular y asimétrica, funcionan como un filtro al reflejar y absorber algunas componentes del sonido. El efecto de filtrado se produce por la interacción del sonido con los pliegues del pabellón auricular, cabeza, torso y hombros, que modifica el contenido espectral del mismo. Esta clave permite localizar fuentes ubicadas en el plano medio (azimut igual a 0°) y varía en función de la posición y distancia de la fuente sonora así como del contenido en frecuencia del estímulo (Blauert 1997).

Estas claves para localizar una fuente sonora son retomadas y explicadas con más detalle en el Capítulo 5, donde se desarrolla el modelo del oyente.

En relación a la percepción de la distancia, se han identificado un conjunto de claves que estarían involucradas (Coleman 1962; Grantham 1995), aunque aún no se conoce en detalle cómo interactúan entre sí (Brown y May 2005):

- **Intensidad:** si una fuente sonora de intensidad constante se aleja del oyente, la disminución en la intensidad del sonido le servirá como clave para percibir su distancia (a menor intensidad se percibe una mayor distancia y viceversa). Esta clave es especialmente relevante cuando se trata de sonidos familiares. Debido a que las fuentes sonoras también pueden variar de intensidad sin modificar su distancia, es necesario contar con otras claves adicionales como la reverberación y la frecuencia.
- **Reverberación:** en espacios cerrados, la proporción de energía directa (temprana) y energía reverberante (tardía) que alcanza al oyente será menor en el caso de una fuente ubicada a distancia lejana y mayor para las fuentes cercanas.
- **Frecuencia:** generalmente las fuentes sonoras ubicadas a distancias lejanas tienen una mayor pérdida de componentes de alta frecuencia por la absorción del aire y los materiales, por esto un sonido con más componentes de alta frecuencia se percibe a una menor distancia.

En el hombre, la habilidad para localizar sonidos es muy buena en el plano horizontal, algo menos eficiente en el plano medio y la precisión con la que se percibe la distancia es regular (Moore, Perazzo, y Braun 1995).

2.6.3 Localización en campo libre

El término localización se refiere a los estudios realizados en campo libre, donde se presentan fuentes sonoras dispuestas según un determinado arreglo espacial,

como, por ejemplo, parlantes apareados (es decir que sus diferencias espectrales han sido corregidas) u objetos sonoros. En este caso el oyente percibe que el sonido proviene desde algún lugar situado en el espacio externo, fuera de él, por lo que se denomina externalización de la imagen sonora. La posición subjetiva o aparente se denomina localización.

El mínimo ángulo audible (minimum audible angle, MAA) es el umbral que indica el cambio más pequeño en la posición de una fuente sonora que puede ser confiablemente discriminado (Mills 1958). El MAA depende tanto de la frecuencia como de la dirección de arribo de la onda sonora.

En el plano horizontal, para estímulos de banda ancha y para tonos de baja frecuencia, el MAA es de 1-2° para posiciones frontales (Mills 1958; Blauert 1997), se incrementa a 8-10° en posiciones laterales extremas ($\pm 90^\circ$) (Kuhn 1987), y decrece nuevamente a 6-7° si la fuente se ubica detrás del oyente (Mills 1958; Perrott 1969). El umbral MAA es mayor para sonidos de medias y altas frecuencias, especialmente para fuentes ubicadas en ángulos que exceden los $\pm 40^\circ$ en el plano horizontal (Mills 1958; 1960; 1972). En el plano medio el MAA es de alrededor de 3-9° para fuentes ubicadas en el hemisferio frontal (Perrott y Saberi 1990; Blauert 1997).

2.6.4 Lateralización

En los experimentos de lateralización, los estímulos se presentan a través de auriculares y tienen la ventaja de que puede ejercerse un mayor control de los parámetros acústicos del estímulo. Su desventaja reside en que, cuando no se utilizan principios de realidad acústica virtual, la persona tiene la impresión de que el sonido proviene desde algún lugar dentro de la cabeza. Esta percepción se denominada internalización de la imagen sonora. La posición donde se ubica la

imagen sonora a lo largo de una línea imaginaria trazada entre los oídos se denomina lateralización de la imagen (Plenge 1974). Un sonido presentado idénticamente a ambos oídos a través de auriculares, llamado estímulo diótico, se lateraliza en el centro de la cabeza. Se puede lograr que la imagen se mueva, por ejemplo al oído derecho, introduciendo un retardo de tiempo a la entrada del oído izquierdo o bien haciendo más intensa la señal del oído derecho. Estos estudios clásicos con auriculares no consideraban el efecto de filtrado natural (Arias 2009).

En tareas de resolución espacial se determina cuál es la diferencia interaural más pequeña que el sujeto es capaz de detectar (just-noticeable difference, jnd). La medición de esta jnd es equivalente a los umbrales de localización obtenidos a través de la técnica del MAA.

2.6.5 Movimientos de cabeza en la localización sonora

Los movimientos de la cabeza son un componente natural e importante del proceso de localización sonora tanto en humanos como en animales (Thurlow, Mangels, y Runge 1967; Knudsen 1981; Perrott, Ambarsoom, y Tucker 1987; Tollin 2004). Son útiles para resolver ambigüedades en la posición percibida de la fuente y contribuyen a maximizar las diferencias binaurales al aumentar o disminuir la ITD e ILD y modificar el efecto de filtrado natural. Se ha demostrado que cuando el participante mueve su cabeza libremente, su rendimiento para localizar sonidos en el plano horizontal y medio mejora (Wallach 1940; Kato et al. 2003).

Perrott et al. (1990), utilizando un sensor primitivo de movimientos de cabeza y trabajando a oscuras, examinaron la capacidad del sujeto para orientar la posición de su cabeza basándose sólo en información acústica. Destacaron que lo más sorprendente fue corroborar que una respuesta motora gruesa (movimiento de

la cabeza) esté tan bien determinada sólo por la información auditiva, especialmente en la región comprendida entre -30° y $+30^\circ$ del plano medio.

En relación a la percepción de distancia los resultados no son concluyentes. Algunos estudios indican que los movimientos de cabeza no influyen sobre el rendimiento en tareas en las que el participante debe determinar la distancia relativa de una fuente sonora (Simpson y Stanton 1973; Litovsky y Clifton 1992; Rosenblum, Wuestefeld, y Anderson 1996), mientras que otros señalan que estos movimientos pueden mejorar el rendimiento (Ashmead, LeRoy, y Odom 1990; Speigle y Loomis 1993).

Por otro lado, cuando se le permite a la persona caminar y moverse libremente hacia la fuente, su rendimiento para localizar sonidos es muy eficiente sobre todo en condiciones adversas, por ejemplo, en situaciones ruidosas o en las que se presentan confusiones delante-detrás. Ashmead, Davis y Northington (1995), evaluaron la habilidad de participantes adultos para juzgar la distancia de una fuente sonora bajo dos condiciones, caminando o desde una posición estática. Los resultados mostraron que el rendimiento de los participantes fue más eficiente en la primera condición. Loomis et al. (1998) realizaron un estudio similar. Evaluaron el rendimiento de participantes con visión normal ocluida en una tarea de localización de fuentes sonoras ubicadas en diferentes posiciones del plano horizontal y a diversas distancias, en condición dinámica (se les permitía caminar hacia la fuente) o estática (reporte verbal). En similar dirección al estudio anterior, sus resultados mostraron que el rendimiento de los participantes fue más consistente cuando podían caminar hacia la fuente sonora que cuando dieron un reporte verbal.

2.6.6 Audición en ambientes reverberantes: efecto precedente

Un aspecto importante en el estudio de la audición espacial, es el efecto de las

reflexiones sobre la habilidad para localizar o discriminar la posición de una fuente sonora. Casi todos los eventos sonoros cotidianos ocurren en lugares donde hay paredes, techos y objetos que reflejan el sonido, es decir, en ambientes reverberantes, donde el hombre también ha demostrado ser muy hábil para localizar sonidos.

Como se sabe, en un ambiente cerrado (reverberante) el sonido proviene de múltiples direcciones y sufre cambios físicos al reflejarse en las superficies de un recinto. A pesar de la presencia de la complejidad de toda esta información sonora, se ha observado que, bajo ciertas circunstancias, la persona escucha un sólo evento auditivo que localiza en la dirección del sonido que le llegó primero, llamado líder. Este fenómeno de audición espacial se denomina *efecto precedente* y ocurre cuando se presentan sonidos similares desde diferentes lugares separados por un breve retardo de tiempo (Wallach, Newman, y Rosenzweig 1949; Yost y Soderquist 1984; Blauert 1997; Litovsky et al. 1999).

En la mayoría de los estudios experimentales sobre efecto precedente, se presentan situaciones muy simplificadas en las que se utiliza sólo un par líder-retardado, con diferentes retardos entre ambos, que simula el sonido original y su reflexión. En un estudio clásico en campo libre sobre efecto precedente, se emplean dos parlantes equidistantes ubicados a la izquierda y la derecha del participante (Figura 2.7a). Un parlante emite el sonido líder y el segundo, después de un retardo variable, emite el sonido retardado que es una réplica del primero. Cuando el sonido líder y retardado se presentan simultáneamente (el retardo es 0 ms), el participante percibe una única imagen fusionada “fantasma” ubicada a la mitad de distancia entre ambas fuentes. En el caso de los estímulos tipo clics, para retardos de entre 1 ms y 5 ms, la fuente sonora se localiza en la posición del sonido líder y

la información direccional del sonido retardado es prácticamente descartada. Para retardos que exceden los 5 ms a 8 ms, la imagen se “parte” en dos y el participante puede localizar separadamente ambos eventos según la posición de las respectivas fuentes sonoras. El retardo para el cual la imagen fusionada se separa en dos se llama umbral del eco y depende fuertemente del tipo de estímulo empleado (umbral clics \approx 5 ms, umbral sonido del habla \approx 35 ms) (Freyman, Clifton, y Litovsky 1991; Blauert 1997) (Figura 2.7b).

El efecto precedente es un mecanismo inconsciente que permitiría al individuo localizar con precisión fuentes sonoras primarias, pues le atribuye un fuerte pesaje al sonido directo y reduce la influencia de la información direccional de los sonidos retardados, evitando imágenes sonoras múltiples.

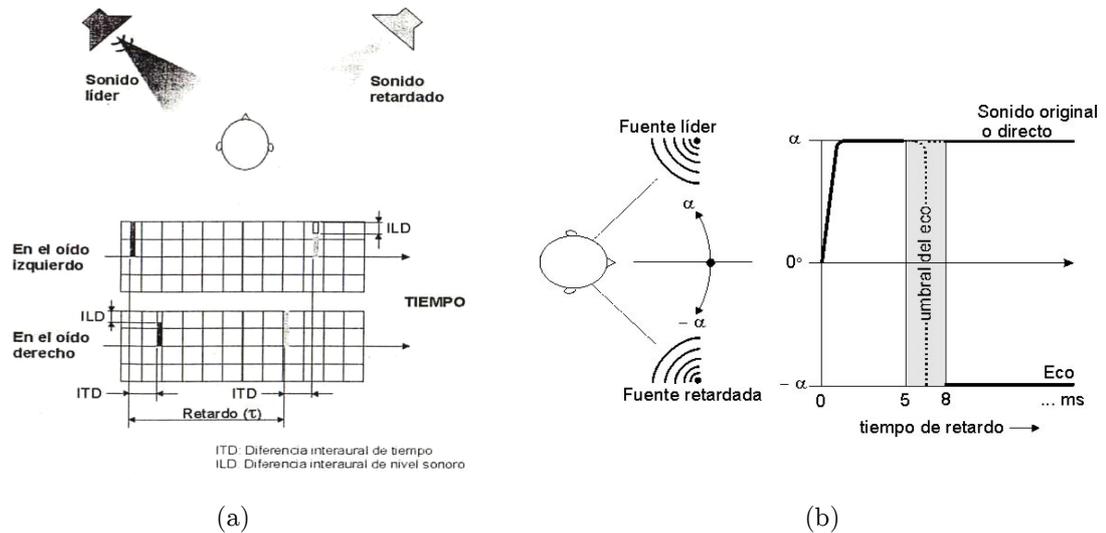


Figura 2.7 (a) Representación esquemática del efecto precedente. Extraído de Arias y Ramos (2003). (b) Efecto precedente. Extraído de Arias (2009).

3 – Fuente sonora

La distribución de presión en la cercanía de una fuente sonora varía según la frecuencia y el espacio. Esta distribución se caracteriza por el diagrama de radiación, conocido como directividad de la fuente. Su modelado es una parte importante (y a menudo olvidada) en los sistemas de realidad acústica virtual. Tradicionalmente se han utilizado grabaciones libres de reflexiones o sonidos sintetizados en reemplazo de la fuente. Tener en cuenta la directividad de la fuente es una mejora sustancial en la sensación de inmersión al entorno (Lentz 2007).

3.1 Señales de entrada al AVRS

Las señales sonoras se pueden dividir en dos grandes categorías: señales grabadas y señales sintetizadas. Las señales grabadas son aquellas que han sido capturadas a partir de una onda sonora real. En esta categoría se incluyen a todas las formas de audio digital (con o sin compresión). En cambio, las señales sintetizadas se crean artificialmente mediante diferentes métodos. También existen métodos híbridos de generación de señales. Esta técnica, en general, involucra a tablas de ondas (de señales ya grabadas), pero en la práctica es poco utilizada.

Savioja et al. (1999) indicaron los requisitos cualitativos generales para las señales de entrada a la fuente sonora en un sistema de acústica virtual. Estos son:

- Las señales deben ser anecoicas, sin presencia de reverberación o propiedades direccionales.
- Las fuentes de sonido deben ser consideradas fuentes puntuales. Una señal estereofónica se puede modelar como dos fuentes puntuales.
- La calidad (relación señal-ruido, cuantización, efectos de la frecuencia de muestreo) de la señal debe ser alta para no degradar la auralización.

3.2 Directividad de una fuente sonora

La mayoría de las fuentes sonoras (altavoces, instrumentos musicales, voz humana, etc.) no irradian la misma energía en todas direcciones, por el contrario concentran mayor energía en algunas direcciones que en otras.

Su modelado exacto es una tarea compleja y requiere de muchos recursos computacionales. Por lo tanto, se utilizan los datos numéricos obtenidos de mediciones de directividad. Con ellas se construyen bases de datos de patrones de directividad para diferentes tipos de fuentes sonoras, tales como instrumentos musicales o la voz humana (Meyer 2009).

Los resultados de las mediciones se suelen representar en los diagramas polares de directividad. En la Figura 3.1 se muestran los diagramas de un oboe para diferentes frecuencias. Es común trazar por separado el plano medio y el plano horizontal. En la Figura 3.2 se muestran los diagramas de directividad tridimensionales para el mismo instrumento.

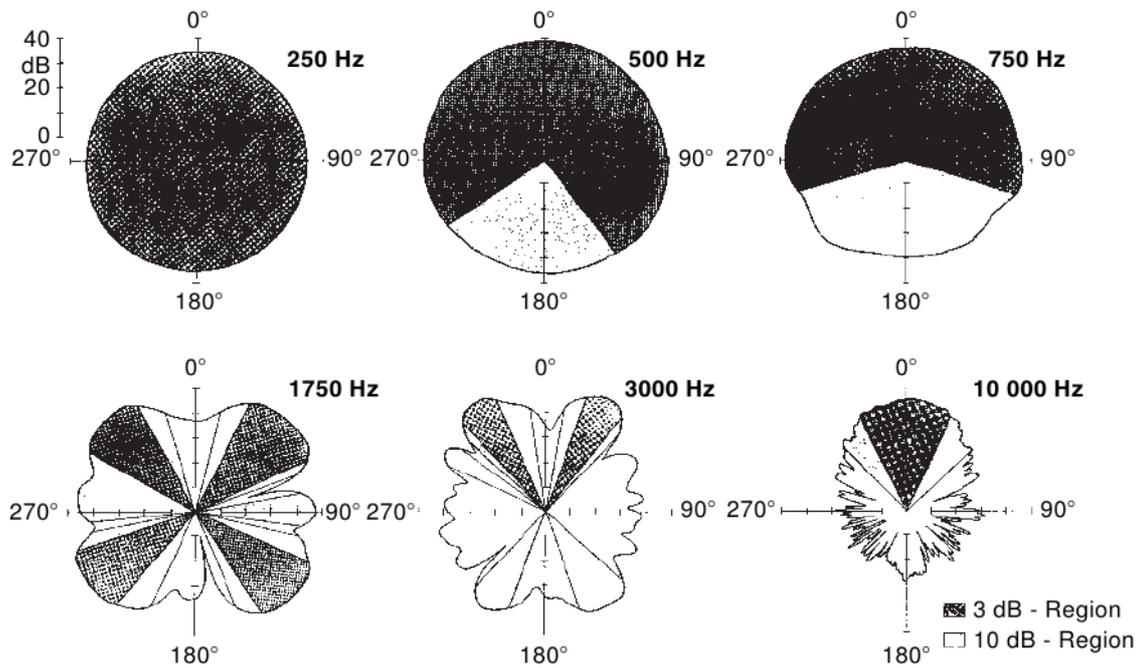


Figura 3.1 Diagramas polares de directividad en el plano horizontal de un oboe para diferentes frecuencias. La dirección 0° corresponde al eje del instrumento. Extraído de Meyer (2009).

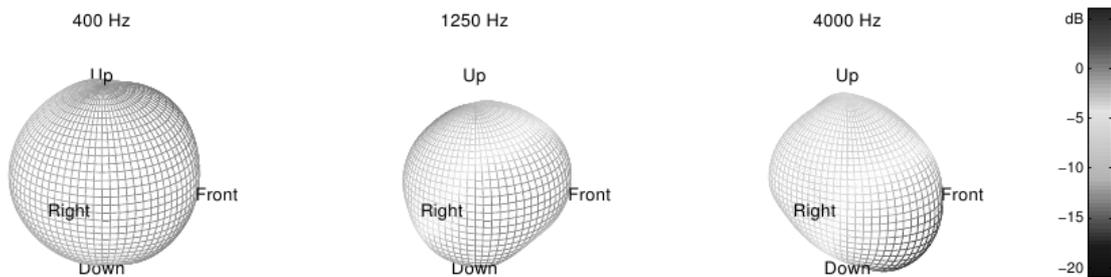


Figura 3.2 Diagramas de directividad tridimensionales de un oboe para diferentes frecuencias.

Extraído de Lentz (2007).

3.2.1 Instrumentos musicales

Medir los patrones de directividad de instrumentos musicales no es una tarea fácil. Las señales emitidas por los instrumentos musicales se deben grabar en un entorno libre de reflexiones⁵ y en campo lejano (Giron 1996). Como el sonido emitido depende del intérprete del instrumento, todas las direcciones deben ser capturadas simultáneamente, requiriendo una gran cantidad de micrófonos para cubrir correctamente sus características direccionales.

3.2.2 Voz humana

El patrón de radiación de la voz humana hablada o cantada es diferente. A su vez, no es constante porque depende de la pronunciación o del texto musical (Vorländer 2007). Existen cabezas artificiales que reproducen los patrones de directividad de la voz hablada y se pueden utilizar para su modelado. Además, Kob (2002) desarrolló un método para medir el patrón de directividad de la voz cantada.

Como ejemplo, en la Figura 3.3, se muestran unos diagramas de directividad tridimensionales para una voz cantada.

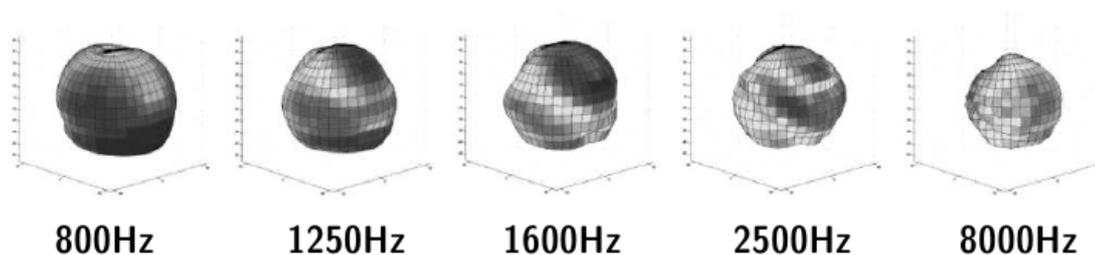


Figura 3.3 Diagramas de directividad tridimensionales de la voz cantada para diferentes frecuencias.

Adaptado de Vorländer (2007).

⁵ En cámara anecoica o en sala muy absorbente, como es un estudio de grabación.

3.3 Modelado de la directividad

La auralización en tiempo real requiere de modelos simplificados que sean eficientes desde el punto de vista de procesamiento de señales y adecuados desde el punto de vista perceptual. El enfoque más simple es modelar la directividad como conos que representan regiones de niveles de presión constantes. Una fuente muy básica se puede simular con este método. Sin embargo, en general, se requiere una representación más exacta, especialmente para instrumentos y para la voz donde su directividad, además, depende de la frecuencia.

Si bien hay diferentes estrategias que se pueden seguir, la más utilizada es la del filtrado direccional. En este enfoque las propiedades de directividad de las fuentes sonoras son modeladas eficientemente junto con los métodos de rayos utilizados para modelar el recinto (ver Capítulo 4). La salida de la fuente sonora $\hat{y}(n)$ alimenta a M filtros $D_{\alpha_i}(z)$ que dependen del ángulo de salida de la reflexión modelada, donde $i = 1, \dots, M$ representa la i -ésima dirección tomada en cuenta en la simulación (Figura 3.4). Generalmente estos filtros son pasa-bajo de bajo orden (cuarto o quinto orden), por lo que satisfacen los requerimientos del tiempo real (Savioja et al. 1999).

Otros modelos, tales como el de “grupo de fuentes elementales” (Savioja et al. 1999) o el método multicanal (Rindel et al. 2004) son menos prácticos para la auralización en tiempo real. Requieren modelar cada fuente sonora como múltiples fuentes elementales. Cada una de estas fuentes produce una señal de salida, las cuales se deben procesar diferenciadamente.

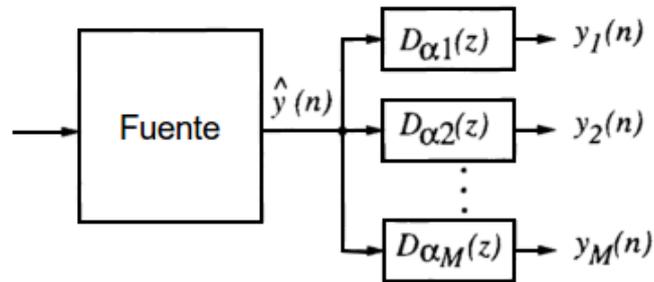


Figura 3.4 Esquema del filtrado direccional

3.4 Fuente sonora omnidireccional

Se define como fuente sonora omnidireccional (omnidirectional sound source, OSS) a la fuente que es capaz de producir igual nivel sonoro en todas las direcciones y frecuencias.

Para que la fuente pueda establecer un campo sonoro difuso a una determinada distancia, se deben cumplir dos propiedades: el isotropismo y la homogeneidad. La primera, establece que la radiación sonora debe tener características esféricas. La segunda, establece que la densidad de energía sonora debe ser igual en cualquier punto de una superficie esférica en cuyo centro se encuentra la fuente.

La omnidireccionalidad puede ser lograda mediante la disposición de 12 altavoces (transductores) en un poliedro regular de 12 caras, llamado dodecaedro (Martín et al. 2007). Para las mediciones en recintos se recomienda el uso de una fuente de sonido omnidireccional. La fuente debe tener un patrón de directividad de acuerdo con el estándar ISO 3382 (2009).

En la Figura 3.5 se muestra la OSS desarrollada en el CINTRA y en la Figura 3.6 su diagrama de directividad, medido en campo libre a 1,5 m, en función del nivel de presión sonora (SPL), cada 30°.



Figura 3.5 Fuente sonora omnidireccional.

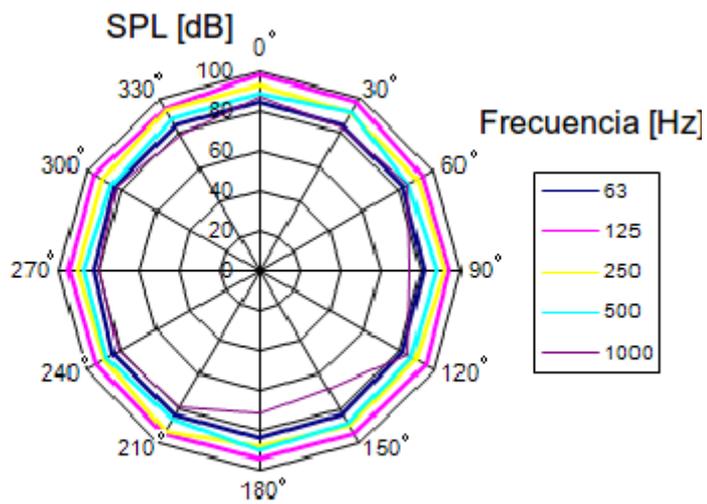


Figura 3.6 Diagrama de directividad de la OSS, en campo libre.

A modo de ejemplo, en la Figura 3.7 se muestra la respuesta impulsiva y la magnitud en el espectro de las frecuencias para la posición frontal, es decir 0° azimut y 0° elevación. La medición fue realizada sobre el eje axial de uno de los transductores.

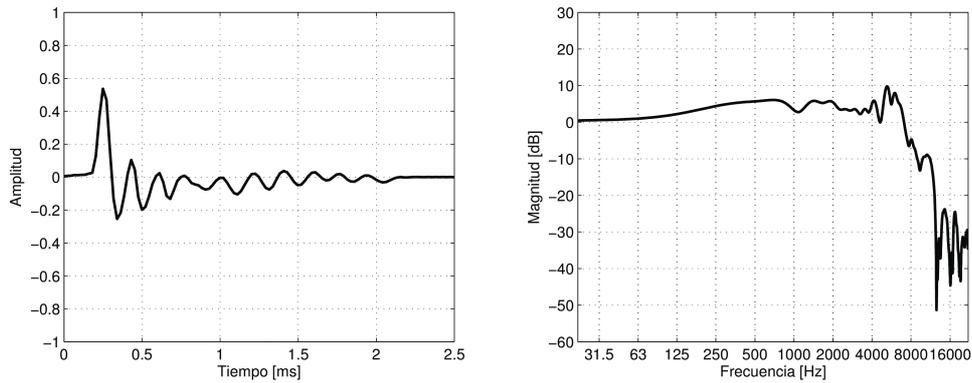


Figura 3.7 Respuesta impulsiva de la OSS y su magnitud espectral

Notar que la fuente tiene la característica que su espectro es relativamente plano hasta los 8 kHz. Con esta fuente se realizaron las mediciones de recintos que fueron simulados por el AVRS (ver Capítulo 7).

4 – 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. Estos cambios dependen de factores tales como el ángulo de incidencia y la relación entre la longitud de onda y el tamaño del objeto, entre otros. Las ondas sonoras emitidas por una fuente, además son alteradas por el medio de propagación (aire en este caso) y por las características de absorción de las paredes y demás objetos del recinto.

4.1 Modelado de la acústica de recintos

En diferentes áreas de la acústica es un tema de interés modelar el campo sonoro producido por una fuente sonora en un espacio cerrado de geometría compleja. En los primeros intentos se utilizaron modelos a escala, que luego fueron relegados a casos muy particulares, tales como el diseño de salas de conciertos (Xiang 1991), dando paso a la simulación acústica por computadora que permite reducir dramáticamente los costos.

El modelado computacional de la acústica de recintos se inició en los años 60 con Schroeder (1961), y luego fue llevado a la práctica por Krokstad et al. (1968) y Schroeder (1973). En los últimos años, con el poder de cálculo de las computadoras

modernas, la simulación acústica de recintos está tomando cada vez más protagonismo.

Estrictamente hablando, modelar el campo sonoro de un recinto consiste en resolver las ecuaciones diferenciales de la onda en el espacio y en el tiempo (ecuaciones de Helmholtz). Hallar una solución analítica es engorroso debido a la dificultad de establecer los límites de integración, por lo que rara vez se realiza. En general, se apela a abordajes con soluciones aproximadas.

Hay tres enfoques diferentes para modelar computacionalmente la acústica de un recinto (Carrión Isbert 2001):

- Métodos basados en la teoría ondulatoria (basados en ondas).
- Métodos basados en la teoría geométrica (basados en rayos).
- Métodos basados en la teoría estadística.

Los métodos de rayos basados en la teoría geométrica presentan limitaciones en el modelado de las bajas frecuencias, lo cual es resuelto por la teoría ondulatoria. La determinación de la frecuencia de transición entre ambos métodos fue propuesta por Schroeder, quien estableció una relación entre el volumen V y el tiempo de reverberación T_{60} del recinto. La denominada frecuencia de Schroeder f_s (Schroeder 1962b) está dada por

$$f_s \approx 2000 \sqrt{\frac{T_{60}}{V}}. \quad (4.1)$$

Por encima de esta frecuencia límite, las numerosas resonancias, con diferentes fases, comienzan a interferir entre sí en un proceso cercano a lo estocástico.

En la Figura 4.1 se muestra un esquema con los diferentes tipos de modelado.

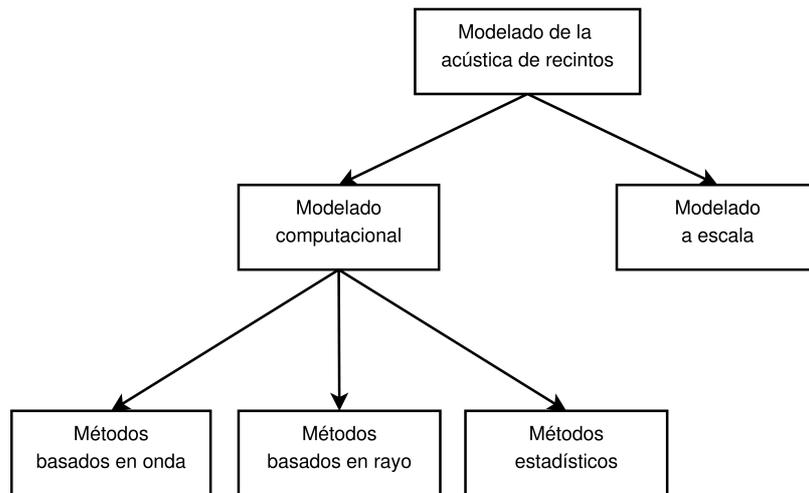


Figura 4.1 Modelos para la acústica de recintos

En los últimos años, han ganado interés los métodos de modelado computacional aplicables a la teoría ondulatoria, tales como el método de elementos finitos (finite element method, FEM), el método de elementos de contorno (boundary element method, BEM) y el FDTD (finite-difference time-domain) (Kuttruff 1995; Botteldooren 1995). Pero la potencia de cálculo que requieren es enorme y su aplicación está restringida a las bajas frecuencias (Rindel 2000).

Los métodos basados en la teoría geométrica más usados son: el trazado de rayos (ray-tracing) (Krokstad, Strom, y Sørnsdal 1968; Kulowski 1985) y sus derivados, el cone-tracing (Van Maercke 1986), el beam-tracing (Funkhouser et al. 2004); y el método de fuente-imagen (image-source method, ISM) (Allen y Berkley 1979; Borish 1984). Utilizan el concepto de rayo sonoro, que se define como una pequeña porción de una onda esférica con un ángulo de apertura infinitesimal que nace desde el centro de la esfera (Kuttruff 2009). El rayo representa la trayectoria de la onda desde una fuente sonora hasta el oyente, y se debe resolver tanto la influencia del medio de propagación como los cambios que se producen cuando

impactan en las superficies.

Los métodos de modelado estadístico, como el análisis de energía estadístico (statistical energy analysis, SEA) (Lyon y DeJong 1995), son usados principalmente en la predicción del aislamiento entre recintos acoplados. No son adecuados para la simulación acústica de recintos ya que no modelan el comportamiento temporal del campo sonoro (Savioja 1999).

El objetivo de estos modelos es obtener la respuesta impulsiva del recinto (RIR), a partir de la cual se calculan diferentes parámetros acústicos⁶, como el tiempo de reverberación (ver Capítulo 2). En este capítulo, se modelará la RIR utilizando un abordaje físico-perceptual de manera que su tiempo de cálculo sea acorde a los requerimientos de un sistema de simulación en tiempo real.

4.2 Geometría del recinto

Para implementar los algoritmos de los modelos, es necesario representar la geometría del recinto como un objeto matemático. El paso inicial es crear un modelo CAD (computer-aided design) del recinto en formato DXF (drawing exchange format)⁷. La cantidad de polígonos de este modelo puede ser muy alta, especialmente cuando se importa de un software de diseño arquitectónico. Puede incluir muchos detalles pequeños, como escalones o los marcos de puertas y ventanas (Figura 4.2).

⁶ Los cuales están definidos en la norma ISO 3382 (2009). Ver Apéndice A para más detalle.

⁷ Formato creado por Autodesk para intercambio de modelos CAD entre diferentes programas.

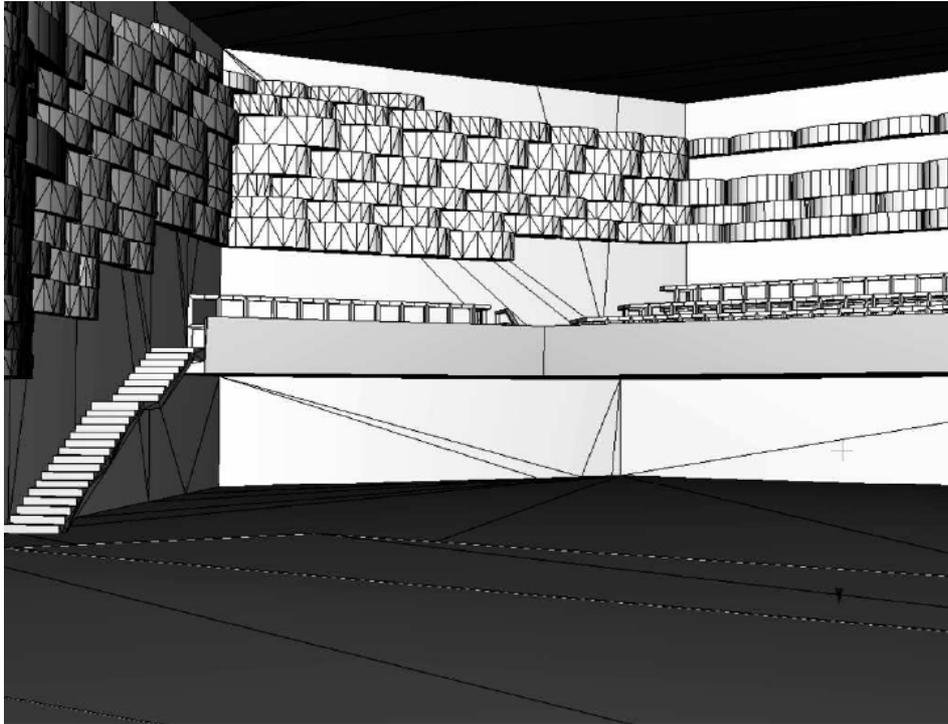


Figura 4.2 Vista del modelo de un recinto con muchos detalles pequeños.

Extraído de Vorländer (2007).

Sin embargo, estos elementos que son esenciales para dar una impresión realista a la vista, son invisibles para las ondas sonoras. Los objetos y los detalles que son pequeños en comparación con las longitudes de onda, deben ser reemplazados por superficies planas con las propiedades acústicas adecuadas (ver sección 4.7 más adelante) (Figura 4.3). Las paredes, puertas y ventanas también son representadas mediante superficies contenidas en planos infinitos.

Como regla general, las superficies de un recinto se deben representar con una resolución de 0,5 m, correspondiente a una longitud de onda de una frecuencia cercana a los 700 Hz (Vorländer 2007).

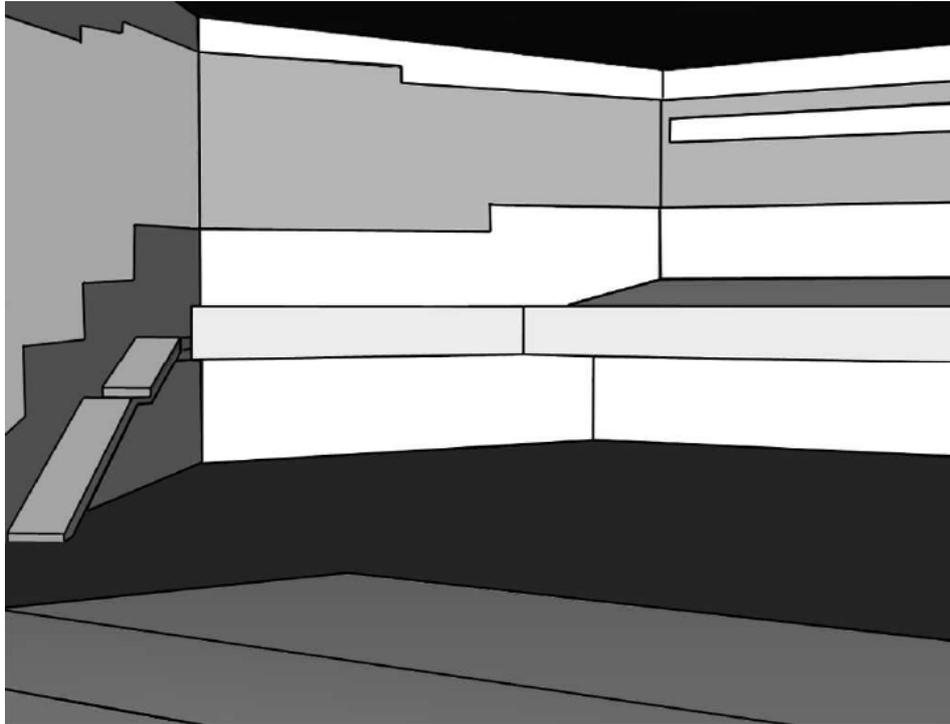


Figura 4.3 Vista del modelo de un recinto simplificado desde el punto de vista acústico. Extraído de Vorländer (2007).

En esta tesis se utiliza un modelado geométrico simplificado. Todas las superficies están representadas por polígonos de cuatro vértices, con un orden de generación en sentido antihorario, como lo muestra la Figura 4.4.

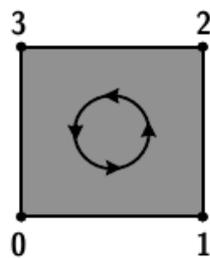


Figura 4.4 Orden de generación de los vértices

Las superficies poseen una cara reflejante y la otra opaca, y aquellas de mayor orden (cilíndricas, esféricas, parabólicas, etc.) son aproximadas mediante superficies planas. Además, es necesario asignarles un material con sus propiedades acústicas. Las superficies complejas, que están conformadas por diferentes materiales, se crean a partir de superficies simples.

4.3 Respuesta impulsiva del recinto

Asumiendo que existe un camino libre de obstáculos entre la fuente y el oyente, la respuesta impulsiva del recinto (RIR) caracteriza, en el dominio del tiempo, la transmisión del sonido entre la fuente y el oyente en dicho recinto. Esta respuesta no es única y depende de la ubicación de la dupla fuente-oyente. Para cada configuración fuente-oyente, existe una RIR que la caracteriza.

Desde el punto de vista del oyente, éste escuchará primero el sonido directo, seguido por las reflexiones tempranas (especulares y difusas) que son producidas en las superficies más cercanas. Luego de algunos milisegundos (entre 50 a 100 ms dependiendo de la geometría de la sala), la densidad de las reflexiones aumenta, y su comportamiento se asemeja a un proceso aleatorio e independiente de la ubicación del oyente. Estas reflexiones se denominan reflexiones tardías o cola de reverberación. En la Figura 4.5 se muestran estas tres partes que constituyen la RIR: sonido directo, reflexiones tempranas y reflexiones tardías.

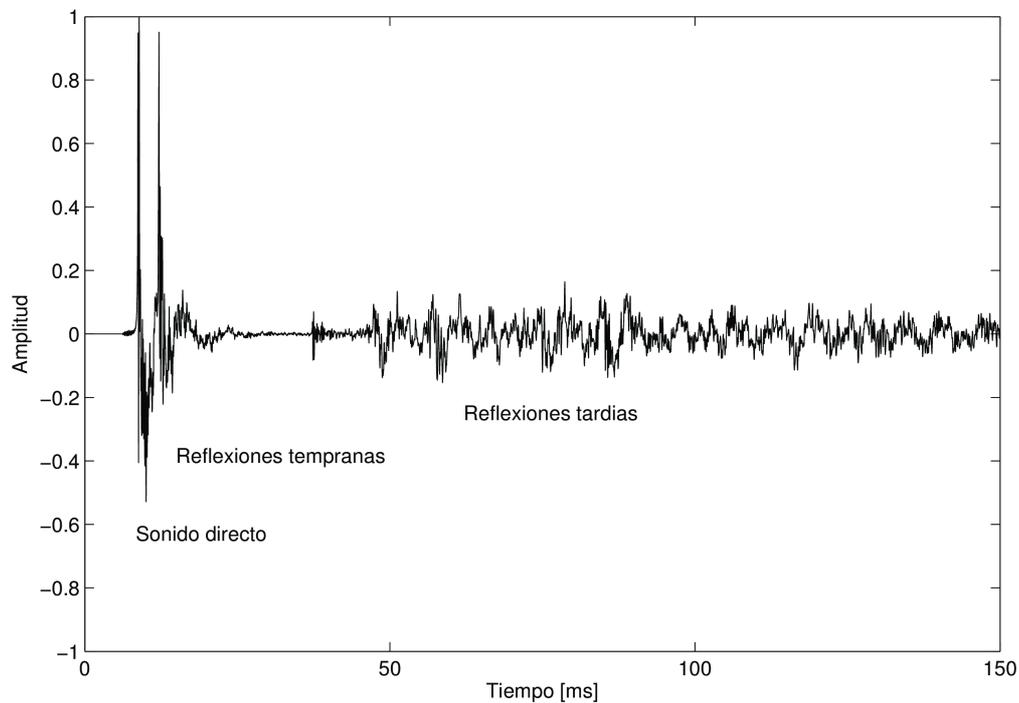


Figura 4.5 Extracto de una RIR para visualizar el sonido directo, las reflexiones tempranas y las reflexiones tardías

Las direcciones de las reflexiones tempranas (junto a la del sonido directo), contribuyen a que una persona pueda localizar con precisión una fuente sonora. Por tal razón, es crucial determinarlas con precisión. Por el contrario, las direcciones de las reflexiones tardías son irrelevantes para localizar una fuente, pero son claves que el ser humano utiliza para percibir las dimensiones, la calidad acústica del recinto y la distancia a la fuente sonora (Blauert 1997).

Numerosas investigaciones se han llevado a cabo con el objetivo de relacionar la RIR con aspectos perceptuales concernientes a la calidad acústica de los recintos (Schroeder, Gottlob, y Siebrasse 1974; Souldre y Bradley 1995; Beranek 1996a; Okano 1998). Aunque los resultados no son aún concluyentes (Morimoto,

Nakagawa, y Iida 2008), se ha postulado que las reflexiones tempranas y tardías contribuyen a determinar lo que se conoce como el ancho aparente de la fuente (apparent source width, ASW) (Keet 1968) y el envolvimiento del oyente (listener envelopment, LEV) (Bradley y Soulodre 1995) respectivamente. Ambos aspectos hacen a lo que se denomina “impresión espacial”⁸.

La norma ISO 3382 describe métodos para medir el tiempo de reverberación en recintos y sugiere otra serie de parámetros calculados a partir de la RIR (ISO 2009). No hay acuerdo sobre los valores óptimos de dichos parámetros salvo para algunos tipos de recintos, según la actividad que en ellos se desarrolla, como es el caso de salas de concierto y casas de ópera (Beranek 1996b). De todos modos, el conocimiento de estos parámetros permite evaluar las propiedades acústicas de un recinto, controlar el campo acústico, aumentar el confort auditivo, elegir los materiales, ubicación de diferentes elementos acústicos (como absorbentes, reflectores y difusores sonoros) (Carrión Isbert 2001; Ferreyra y Elías 2006).

En los sistemas en tiempo real las limitaciones que impone la capacidad computacional obliga a realizar simplificaciones. Como a cada parte de la RIR se le atribuyen diferentes implicancias perceptuales, se utilizó una metodología híbrida para su construcción. El sonido directo y las reflexiones tempranas especulares se simulan mediante un método basado en rayos: el método de la fuente-imagen. Para las reflexiones tempranas difusas se utilizan filtros de difusión de Pellegrini y para las reflexiones tardías una red de retardo realimentada, que es una estructura de filtros recursivos. A continuación se desarrollarán dichos métodos.

8 Imagen conceptual que un sujeto obtiene espontáneamente de las características (tipo, tamaño y propiedades) del recinto (Blauert 1997).

4.4 Reflexiones especulares

La ley de la reflexión total, conocida como reflexión especular, fue introducida por W. Snel van Royen en el siglo XV y establece que el ángulo del rayo reflejado es igual al incidente con respecto a la normal a la superficie que la originó. Su aplicabilidad se restringe a casos en que la longitud de la onda es pequeña con respecto al tamaño de las superficies y grande en relación a sus irregularidades y curvaturas.

4.4.1 Método de la fuente-imagen

El método de la fuente-imagen (ISM) permite determinar con precisión el camino y dirección de los rayos sonoros que llegan al oyente. Pero tiene como desventaja que su costo computacional crece exponencialmente a medida que aumenta la cantidad de reflexiones a calcular. Por esto, sólo se usa para generar las reflexiones tempranas especulares que, como se dijo, su relevancia perceptual es crucial.

El principio básico es sencillo: los rayos sonoros reflejados en una superficie son reemplazados por rayos directos provenientes de fuentes imágenes espejo reflejadas de la fuente real (Allen y Berkley 1979). Como se puede observar en la Figura 4.6, una fuente imagen o fuente virtual (virtual source, VS) S_1 se encuentra equidistante de la fuente real S_0 según la normal a la superficie que le da origen.

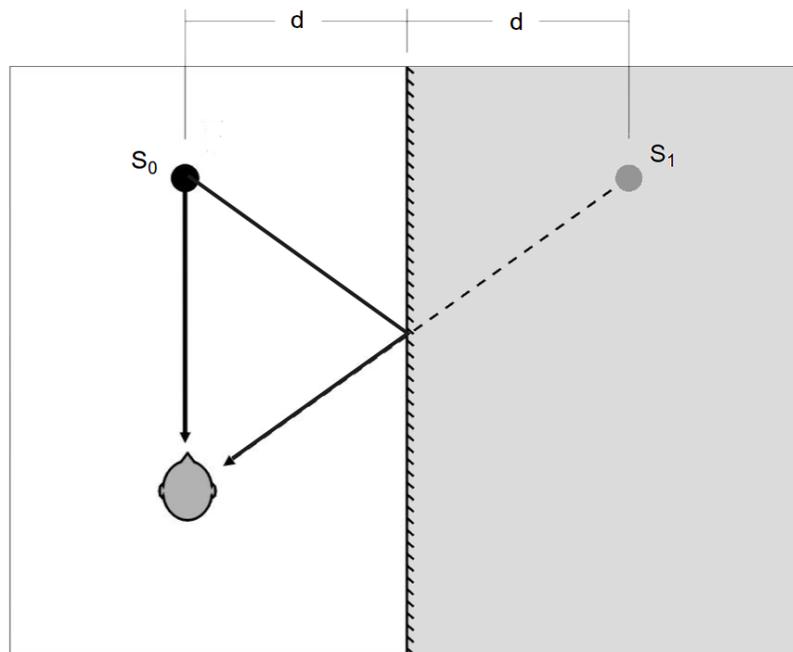


Figura 4.6 La distancia recorrida y la dirección de arribo del rayo que se refleja en la pared son iguales a las del rayo emitido por la VS

En un recinto dado, el campo sonoro total en la posición del oyente L está conformado por el rayo directo, originado por la fuente real S_0 (el que recorre el mínimo camino entre la fuente y el oyente), y por el conjunto de rayos originados en las VSs (S_1, S_2, \dots, S_n). En la Figura 4.7 se muestra un ejemplo en dos dimensiones de los rayos sonoros que llegan al oyente luego de que éstos se reflejen en una superficie.

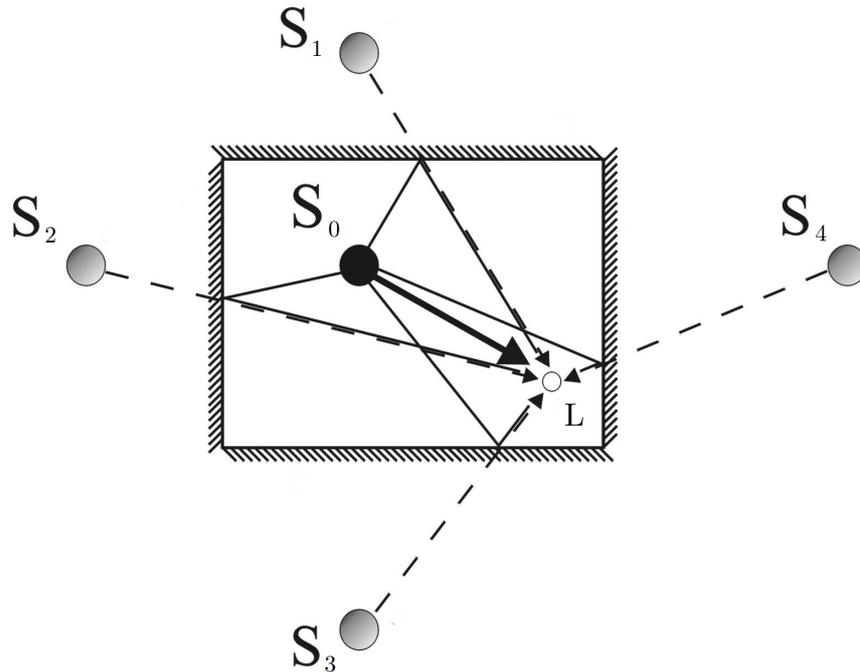


Figura 4.7 Los rayos reales (que se reflejan en superficies) se reemplazan por los rayos que emiten las VSs

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

4.4.2 Cálculo de la posición de una VS

La ubicación en el espacio de un plano que contiene a una superficie reflejante está determinada por dos parámetros: la normal unitaria al plano \hat{n} y su distancia p al origen de coordenadas del recinto.

La posición en el espacio \mathbf{s}_v de una VS se encuentra haciendo

$$\mathbf{s}_v = \mathbf{s}_r + 2d\hat{n} \quad (4.2)$$

siendo

$$d = p - \mathbf{s}_r \hat{\mathbf{n}} \quad (4.3)$$

donde d es la distancia de la VS al plano; y \mathbf{s}_r las coordenadas de la posición en el espacio de la fuente real (Figura 4.8).

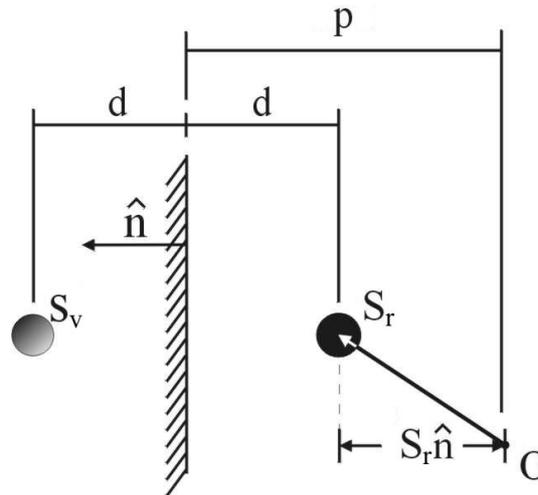


Figura 4.8 Cálculo de la posición de una VS

4.4.3 Generación de las reflexiones tempranas especulares

El procedimiento consiste en generar las VSs, reflejando la fuente real a través de cada superficie reflejante. Luego, cada VS calculada se refleja nuevamente en cada superficie, y así sucesivamente. Como se puede apreciar el algoritmo es recursivo. De esta manera, las VSs pueden ser organizadas en una estructura de datos tipo árbol n -ario (Borish 1984).

Si el recinto tiene n superficies, entonces hay n posibles VSs de primer orden (S_1, S_2, \dots, S_n). A su vez, cada una de éstas (VSs padres o progenitoras) da origen a $(n - 1)$ VSs de orden superior (VSs hijas). Para un orden dado i , el número de VSs crece exponencialmente de la forma $n(n - 1)^{i - 1}$ (Tsingos y Gascuel 1997). Como se habrá podido advertir, el orden de una VS está dado por la cantidad de

veces que el rayo sonoro se reflejó en las diferentes superficies antes de llegar al oyente.

La distancia de cada VS al oyente es igual al camino recorrido por el rayo reflejado real y su dirección resulta de unir la VS con el oyente. Las características de cada VS dependen de las características de la fuente real y de las propiedades físico-acústicas de los materiales en los que se reflejó y el medio de transmisión (aire) (ver secciones 4.6 y 4.7 más adelante).

Este método exacto tiene solución trivial para un recinto con forma de cuboide. Pero para recintos con formas arbitrarias, puede suceder que algunas VSs de orden superior no contribuyan al campo sonoro total (Borish 1984). Como en esta tesis es de interés generalizar a recintos de formas arbitrarias, se deben realizar pruebas adicionales de validez, proximidad y audibilidad (también llamada visibilidad).

Validez

Considerando el lado reflejante de las superficies como si fueran espejos, una VS no es válida si fue creada por una cara no reflejante. Las reflexiones no válidas pueden ser detectadas si la distancia calculada en la ecuación (4.3) es negativa, lo que implicaría que la VS está dentro del recinto.

Proximidad

Es la prueba que indica la finalización del algoritmo recursivo. Existen dos criterios para terminar la generación de VSs.

Una forma es especificar la longitud máxima que pueden recorrer los rayos. Las VSs que están más lejos del oyente que este valor, se descartan. Así la línea de descendencia en el árbol de dicha VS finaliza, ya que es imposible que una VS hija

esté más cerca del oyente que la VS padre, porque que el camino que debe recorrer el sonido será mayor. Es una buena pauta descartar las VSs que se encuentren a una distancia tal que llegan al oyente en un tiempo posterior a la finalización de las reflexiones tempranas. Por ejemplo, para un recinto donde las reflexiones tempranas llegan hasta los 100 ms, la distancia máxima establecida será de 34,3 m si la velocidad del sonido es de 343 m/s.

La otra forma de terminar el algoritmo es especificar un orden máximo de las VSs. Esto se realiza, principalmente, para ahorrar costo computacional y se determina en función de la capacidad de cómputo de la máquina donde se ejecuta la simulación. Lógicamente, a mayor orden, mayor precisión.

En la tesis se utilizó una combinación de los dos métodos, es decir, cada línea de descendencia de las VSs en el árbol se termina por aquel criterio que se cumple primero.

Audibilidad

Con esta prueba se busca asegurar que la VS hija sea “visible” (audible) para el oyente a través de la superficie por la que fue creada. La VS es audible, si la recta trazada entre ésta y el oyente intercepta al plano de la superficie reflejante, en un punto que está dentro de los límites de la superficie (Figura 4.9).

Hay diferentes métodos para dar solución a este problema, conocido como el problema del punto en el polígono (point-in-polygon) (Haines 1994). El método elegido en esta tesis es el de crear vectores desde el punto de intersección hasta cada uno de los vértices de la superficie. Los productos cruz de los sucesivos pares de estos vectores son vectores ortogonales al plano de dicha superficie. Si cada uno de los vectores ortogonales calculados tienen la misma dirección, entonces el punto

de intersección está dentro del polígono (Figura 4.9a), y la VS es visible. Por el contrario, si algún vector ortogonal tiene dirección opuesta ($V_2 \times V_3$ en el ejemplo), el punto de intersección se encuentra fuera y la VS no es visible (Figura 4.9b).

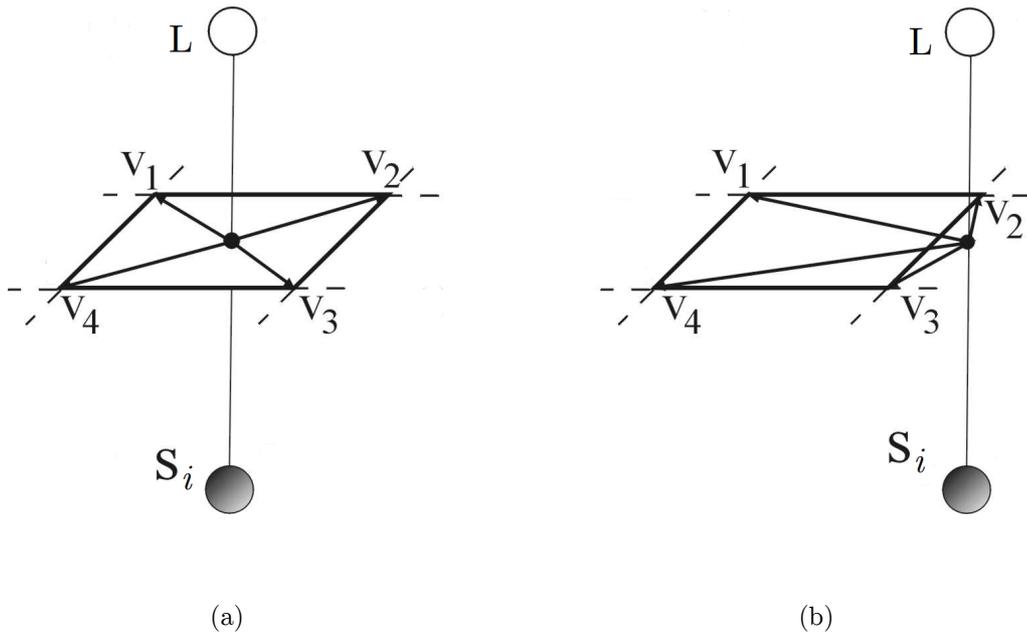


Figura 4.9 Determinación si el punto pertenece al polígono. En (a) el punto pertenece a la superficie evaluada; en (b) el punto no pertenece a la superficie.

Para las reflexiones de primer orden, esta verificación es suficiente. Pero para reflexiones de orden superior se deben realizar pruebas adicionales de audibilidad. Borish (1984) indica que los caminos que involucran más de una reflexión requieren del cálculo de oyentes virtuales (oyentes imágenes) de orden superior que apuntan hacia la correspondiente VS. De esta forma se valida cada segmento del camino de sonido real. Sin embargo, en lugar de calcular las posiciones de los oyentes virtuales, se propone otro procedimiento que tiene un menor costo computacional (Tommasini y Ramos 2008), que se explica a continuación.

La solución consiste en reconstruir la trayectoria del rayo desde el oyente a la fuente y verificar que los puntos de intersección con los planos pertenezcan a las superficies que originaron las VSs. Si todos los puntos de intersección están dentro de los límites de las superficies respectivas, entonces la VS cumple con el criterio de audibilidad de orden superior. Si, en cambio, durante el procedimiento algún punto de intersección no está dentro de los respectivos límites, se aborta el procedimiento, y la VS no cumple con este criterio. Este procedimiento se repite con todas las VSs de orden superior.

Por ejemplo, en la Figura 4.10 se desea verificar si la VS de orden superior S_2 cumple con el criterio de audibilidad. Se calcula el punto intersección p_1 de la línea que conecta al oyente L y S_2 con el plano c que contiene a la superficie que la generó. Como p_1 se encuentra dentro de los límites de la superficie (línea negra gruesa), se sigue verificando con la VS padre S_1 . Ahora se calcula el punto de intersección p_2 de la línea que conecta al punto p_1 (recién calculado⁹) y S_1 con el plano a que la originó. Como p_2 no está dentro de los límites de la superficie, se concluye el procedimiento y S_2 no cumple el criterio de audibilidad de orden superior.

⁹ Aquí, por ejemplo, se evita el cálculo adicional de oyentes virtuales.

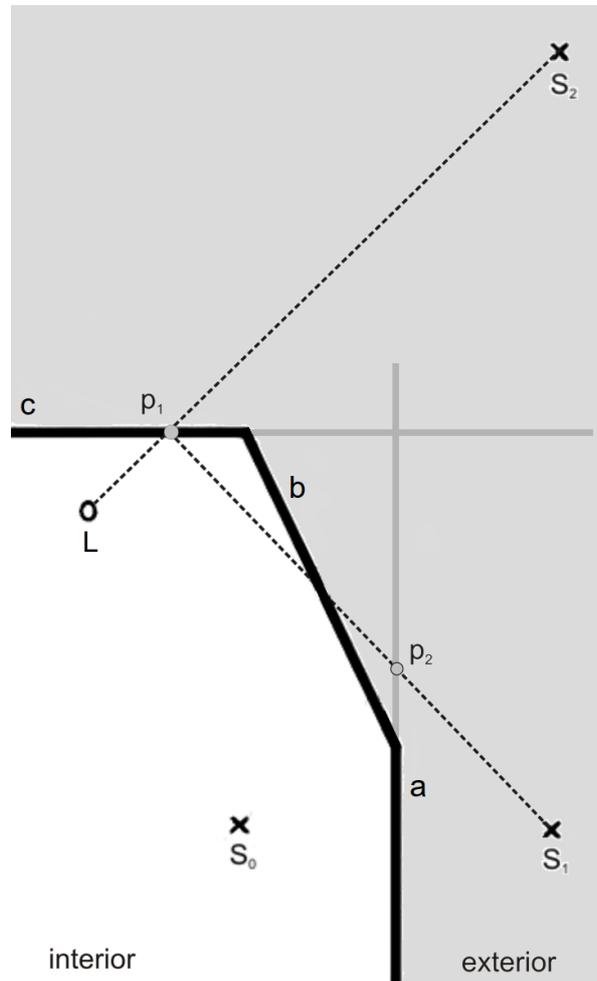


Figura 4.10 Prueba de audibilidad de orden superior

4.4.4 Distribución temporal del campo sonoro en el oyente

Hay que tener en cuenta que sólo las VSs que son audibles contribuyen al campo sonoro en el oyente. Sin embargo, las VSs que no son audibles no se descartan, ya que alguna de sus VSs hijas puede serlo. Una VS se descarta sólo cuando no cumple con las pruebas de validez o proximidad.

Supongamos que la fuente real y todas las VSs emiten un impulso discreto de

igual amplitud y al mismo tiempo. Las ondas producidas arribarán al oyente con un cierto retardo de tiempo según la distancia que recorren. La presión sonora en el oyente es inversamente proporcional a la distancia que recorren e igual a $1/ct$, siendo c la velocidad del sonido.

A partir de la distribución temporal del sonido directo y de las reflexiones tempranas especulares calculadas, se puede construir una respuesta impulsiva (también llamada ecograma) (Figura 4.11).

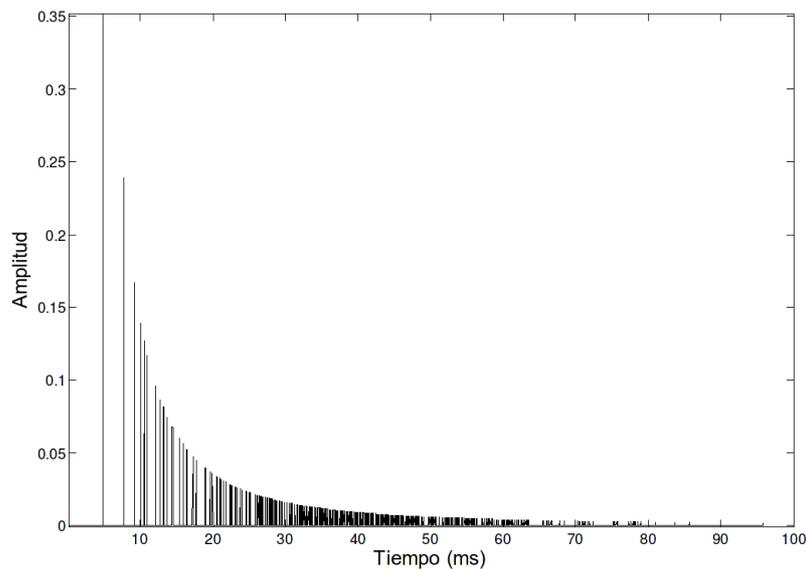


Figura 4.11 Las VSs contribuyen con un impulso discreto retardado y atenuado según la distancia al oyente

Según Kuttruff (2009) la cantidad de rayos (reflexiones) que alcanzan al oyente al cabo de un tiempo t , en un recinto de volumen V , es igual a

$$N = \frac{4\pi c^3 t^3}{3V}. \quad (4.4)$$

La dependencia cúbica con el tiempo hace que la distribución o densidad

temporal de los rayos no sea uniforme. Al comienzo, la densidad de las reflexiones es baja y los valores de energía son significativos. Con el transcurso del tiempo, la densidad aumenta y la energía disminuye. Existirá un instante de tiempo en que el ser humano no podrá distinguir entre dos reflexiones contiguas y su percepción se asemejará a un ruido continuo que se atenúa progresivamente hasta hacerse imperceptible. Es donde comienzan las reflexiones tardías.

4.5 Reflexiones difusas

Cuando un rayo sonoro alcanza una superficie rugosa e irregular, junto con la reflexión especular se originan reflexiones difusas. La difusión está relacionada a la redirección de una porción de la energía especular en direcciones aleatorias no especulares (Figura 4.12). El término reflexión difusa a menudo es utilizado como un equivalente a difusión (Dalenbäck, Kleiner, y Svensson 1993).

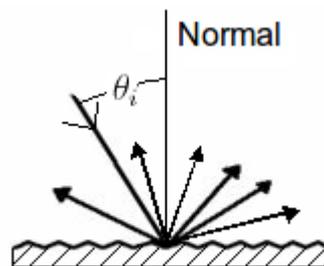


Figura 4.12 Reflexiones
difusas

A modo de ejemplo, en la Figura 4.13 se puede ver gráficamente la reflexión difusa que produce un impulso.

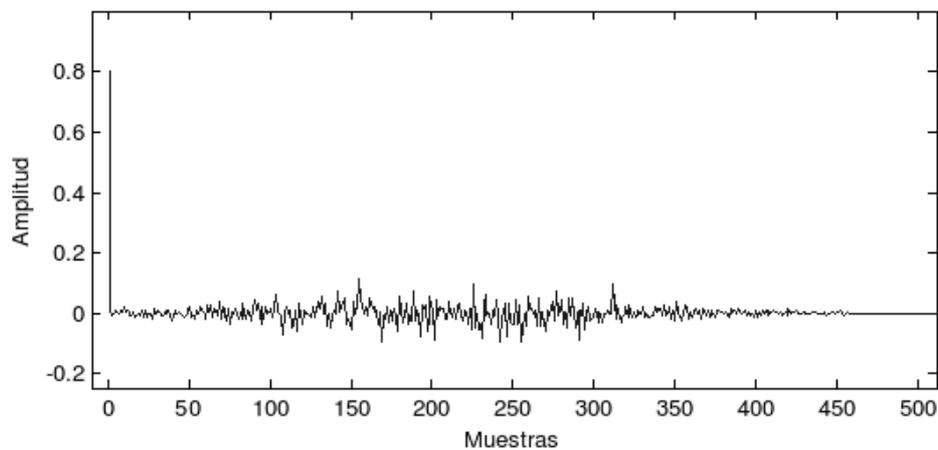


Figura 4.13 Reflexión difusa producida a partir de un impulso

La difusión se produce cuando la longitud de onda es comparable con la rugosidad y asperezas de los materiales. Como estas reflexiones son difíciles de modelar con exactitud, en general se realizan simplificaciones aceptables desde la perspectiva perceptual.

Existen varios métodos para modelar este fenómeno: método de Lambert y Lambert oblicuo (Christensen y Rindel 2005a; Christensen y Rindel 2005b), difracción de bordes basado en la técnica de Biot-Tolstoy-Medwin (BTM) (Torres, Svensson, y Kleiner 2001; Torres, de Rycker, y Kleiner 2004), modelo de “boss” para dispersión de superficies (Chu y Stanton 1990) y el modelo de reflexiones difusas de Pellegrini (Pellegrini 2002). El modelo más satisfactorio para esta tesis, desde el punto de vista perceptual y de su implementación en tiempo real, es el modelo de Pellegrini.

4.5.1 Modelo de reflexiones difusas de Pellegrini

Schroeder (1962a) propone la utilización de filtros pasa-todo (all-pass) para simular la difusión, pero tenían el problema de producir un sonido característico no

deseado, conocido como “coloración”.

Para modelar el comportamiento de propagación en todas las direcciones de manera estocástica, Pellegrini propone utilizar los filtros pasa-todo de Schroeder en cascada con una modificación. La cual consiste en utilizar filtros pasa-todo complementarios para atenuar la coloración.

Partiendo de la función de transferencia del filtro pasa-todo

$$H_1(z) = \frac{-\alpha + z^{-k}}{1 - \alpha z^{-k}} \quad (4.5)$$

donde k es la cantidad de polos y ceros, y α es la ganancia. Se propone un filtro complementario con un retardo adicional m (múltiplo de k) y un factor de atenuación β

$$H_2(z) = \beta z^{-m} \frac{-\alpha + z^{-k}}{1 - \alpha z^{-k}}. \quad (4.6)$$

Realizando la diferencia entre las ecuaciones (4.5) y (4.6) se obtiene la función de transferencia

$$G(z) = \frac{-\alpha + z^{-k} + \alpha\beta z^{-m} - \beta z^{-(m+k)}}{1 - \alpha z^{-k}} \quad (4.7)$$

siendo $\alpha = e^{k\gamma}$, con $0 < \alpha < 1$, y

$$\beta = \frac{1 - \alpha^2}{\alpha} e^{m\gamma} \quad (4.8)$$

$$\gamma = \frac{\ln\left(\sqrt{\frac{1}{2}}\right)}{k + 1} \quad (4.9)$$

donde β es la ganancia de la envolvente del filtro pasa-todo de (4.5) para la m -ésima muestra, y γ indica el factor de atenuación óptimo. Utilizando un valor de $k = 5$ y $m = 10k$, se obtiene la respuesta impulsiva del filtro $G(z)$ mostrada en la Figura 4.14.

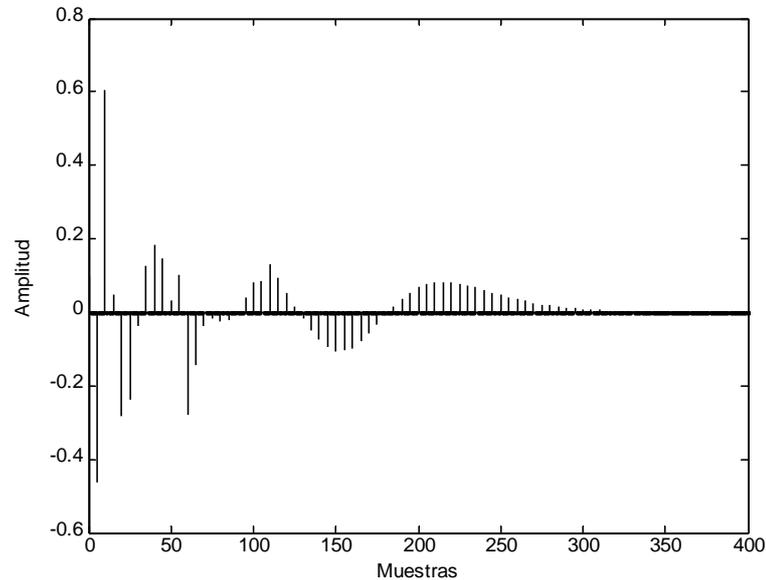


Figura 4.14 Respuesta impulsiva del filtro $G(z)$ (sin sobremuestreo)

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 puede utilizar k para determinar la longitud y luego realizar un sobremuestreo de la respuesta impulsiva para lograr una mayor densidad (Tommasini, Ramos, y Ferreyra 2008a). Se conectaron ocho filtros $G(z)$ en cascada para lograr el filtro de difusión.

4.6 Caracterización del medio de transmisión

Cuando una onda sonora viaja a través de un medio de transmisión (el aire en este caso), éste absorbe parte de su energía. Esta absorción depende de la frecuencia, la distancia recorrida, la temperatura y la humedad del ambiente.

A partir de las ecuaciones estandarizadas de absorción del sonido en el aire (ISO 1993), se calcularon un conjunto de filtros para varios valores de distancia,

para una temperatura de 20°C y una humedad del 20% (Bass, Bauer, y Evans 1972) (Figura 4.15). Con filtros IIR pasa-bajo de segundo orden se puede ajustar su respuesta en frecuencia (Savioja et al. 1999). El efecto de la atenuación por la distancia recorrida se logra multiplicando la salida de los filtros por la inversa de tal distancia.

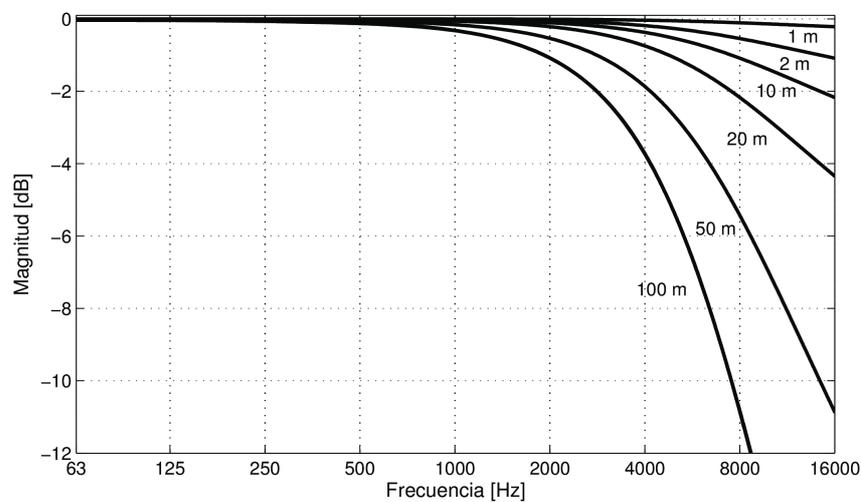


Figura 4.15 Magnitud de los filtros de absorción de aire, sin tener en cuenta la atenuación por la distancia

4.7 Caracterización de los materiales

Cuando las ondas sonoras interactúan con los materiales de las superficies, no toda la energía es reflejada. Parte de ella se absorbe o se transmite al otro lado de la superficie (Figura 4.1). Por acción de los materiales, el rayo reflejado tiene un contenido espectral diferente al del rayo incidente.

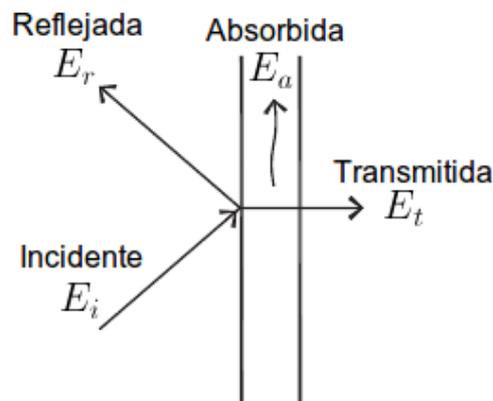


Figura 4.16 Comportamiento de las ondas sonoras cuando inciden sobre una superficie

Este balance de energía se puede expresar como

$$E_i = E_r + E_a + E_t. \quad (4.10)$$

El cociente entre la energía que se refleja y la energía incidente se conoce como coeficiente de reflexión. Es una magnitud compleja característica del material y depende del ángulo de incidencia y de la frecuencia. En acústica arquitectónica es común disponer de los coeficientes de absorción, para frecuencias discretas entre 125 y 4000 Hz espaciadas por bandas de una octava. Para obtener estos coeficientes se utilizan dos métodos tradicionales que son: el método de tubo de onda estacionaria (Ginn 1978; ISO 1996) y el método de medición en cámara reverberante (Ginn 1978; ISO 2003). En general, los coeficientes encontrados en la literatura son independientes del ángulo de incidencia.

Como se vio en el Capítulo 2, la relación que existe entre el coeficiente de absorción $\alpha(\omega)$ y el de reflexión $R(j\omega)$ está dada por

$$\alpha(\omega) = 1 - |R(j\omega)|^2. \quad (4.11)$$

Estos coeficientes de absorción (medidos en sabins) tienen la limitación que en recintos con tiempo de reverberación inferior a 2 s pueden alcanzar valores por

encima de 1 (hasta 1,57 para algunos materiales), lo que resulta en una incoherencia a nivel físico. Norris y Eyring propusieron el uso de la siguiente fórmula para corregir los coeficientes

$$\alpha_{\text{Norris-Eyring}} = 1 - e^{-\alpha_{\text{Sabine}}} . \quad (4.12)$$

A partir de los coeficientes de reflexión se diseñaron filtros IIR de cuarto orden para modelar la absorción de los materiales (Savioja et al. 1999). Éstos fueron interpolados linealmente para frecuencias intermedias (obteniendo valores igualmente espaciados en el espectro) y extrapolados para frecuencias superiores a 4000 Hz (Figura 4.17).

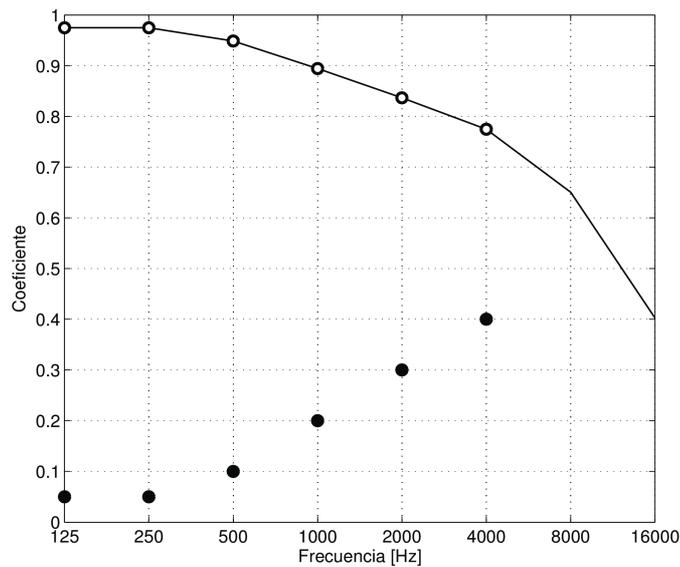


Figura 4.17 Coeficientes para un piso de alfombra de 1/8". Los puntos negros son los coeficientes de absorción α . Los puntos blancos, los coeficientes de reflexión R correspondientes. La línea continua se obtiene por interpolación lineal y extrapolación.

La validez de la simulación de los materiales también se restringe a casos en que la longitud de la onda es pequeña con respecto al tamaño de las superficies y

su radio de curvatura, y grande en relación a las irregularidades y porosidades de los materiales.

El sistema auditivo tiene la particularidad de que la resolución en frecuencia no es uniforme a lo largo del rango audible; es más selectivo en bajas frecuencias y empeora en altas frecuencias. Moore y Glasberg (1996) introdujeron la siguiente expresión que relaciona el ancho de banda con la frecuencia central, denominándola banda rectangular equivalente (equivalent rectangular bandwidth, ERB):

$$\text{ERB}(f) = 0,108f + 24,7 \quad (4.13)$$

donde f es la frecuencia central en hertz, en el rango de 100 Hz a 10 kHz.

En el diseño de los filtros de los materiales, se utiliza la inversa de la ecuación (4.13) como función de ponderación al minimizar el error de ajuste.

4.8 Reflexiones tardías

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. Los algoritmos de reverberación artificial intentan recrear esto usando diferentes técnicas, pero variando sus requerimientos computacionales (Schroeder 1962a; Rubak y Johansen 1999; Lokki y Hiipakka 2001).

Por las razones perceptuales ya vistas, no es necesario modelar las reflexiones tardías de manera precisa como se hace con las tempranas. El grado de precisión difiere de una aplicación a otra y del enfoque a utilizar. El propósito es reproducir sólo las claves consideradas importantes de una reverberación natural. Las reflexiones tardías están relacionadas con aspectos perceptuales particulares, como

por ejemplo, la percepción de la espacialidad y la distancia percibida a la fuente, entre otros.

Dos aspectos perceptuales deben tenerse en cuenta en el diseño de un reverberador artificial: la densidad de tiempo y la densidad de frecuencia. Se definen como la cantidad de eventos temporales por segundo y como la cantidad de componentes en frecuencia por hertz que una persona puede discriminar, respectivamente.

4.8.1 Reverberador de Schroeder

El primer reverberador artificial de señales digitales fue construido por Schroeder (1962a). El diseño original proponía conectar filtros pasa-todo en serie. Luego, el mismo Schroeder propone conectar filtros peine (comb) en paralelo, mejorando el rendimiento. Este diseño es conocido como el reverberador de Schroeder.

El problema de este reverberador es que, para obtener una adecuada densidad en el dominio del tiempo y en el de la frecuencia, se requieren una cantidad importante de filtros, haciendo al sistema ineficiente (Tommasini, Ramos, y Ferreyra 2008a).

4.8.2 Red de retardo realimentada

La red de retardo realimentada (FDN) tiene las ventajas de los filtros peine en paralelo y los filtros pasa-todo en serie en una misma estructura. Se logran densidades (tanto temporales como de frecuencia) aceptables y se evita la percepción de un sonido metálico.

Gerzon (1976) fue el primero en generalizar el concepto a una red multicanal unitaria, equivalente a N filtros pasa-todo unitarios. Posteriormente, Stautner y

Puckette (1982) propusieron una red general de cuatro canales con cuatro líneas de retardo y una matriz de realimentación. A principios de los años 90, Jot (Jot y Chaigne 1991; Jot 1992) introdujo una red con un canal de entrada y uno de salida con N líneas de retardo, generalizando la de Stautner y Puckette. Esta red es conocida como FDN de Jot (Figura 4.18).

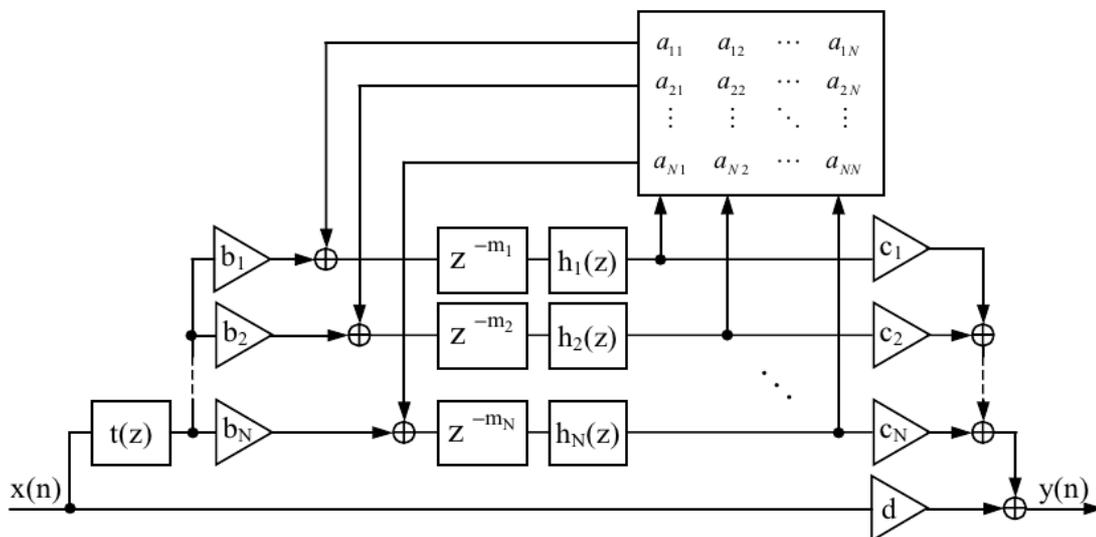


Figura 4.18 FDN propuesta por Jot. Extraída de Frenette (2000).

El uso de una FDN para el modelado de reverberación tardía se justifica en el marco de la teoría estadística de reverberación (Kuttruff 2009) y se basa en la condición de que se logra un solapamiento (o densidad) suficiente de los modos acústicos en el dominio de la frecuencia y de las reflexiones en el dominio del tiempo (Schroeder 1962a; Jot, Cerveau, y Warusfel 1997). Bajo este supuesto, las reflexiones tardías pueden ser modeladas como una señal gaussiana aleatoria con decaimiento exponencial, que se caracteriza por una envolvente de energía

espectral, que se denota $E(\omega)$, y el tiempo de reverberación versus frecuencia, denotado $T_{60}(\omega)$.

La densidad de tiempo D_t y la densidad de frecuencia D_f se pueden calcular de la siguiente manera

$$D_t = \sum_{i=1}^N \frac{1}{\tau_i} \approx \frac{N}{\tau} \quad (4.14)$$

$$D_f = \sum_{i=1}^N \tau_i \approx N\tau \quad (4.15)$$

donde N es la cantidad de líneas, τ_i es el retardo de la i -ésima línea en segundos y τ es el promedio de los retardos de la N líneas.

4.8.3 Matriz de realimentación y retardos

Para que la salida se asemeje a un ruido blanco, es importante la elección de la matriz de realimentación $\mathbf{A} = [a_{ij}]_{N \times N}$ y de los retardos m_i . Cada coeficiente a_{ij} determina la proporción de señal generada por la línea de retardo j que realimenta la línea de retardo i . Jot y Chaigne (1991) proponen la utilización de una matriz “sin pérdidas” de Householder, que tiene la forma

$$\mathbf{A} = \frac{2}{N} \mathbf{u}\mathbf{u}^T - \mathbf{I} \quad (4.16)$$

donde N es la cantidad de líneas de retardo, $\mathbf{u}^T = [1, 1, \dots, 1]$, e \mathbf{I} es una matriz identidad de orden $N \times N$.

La elección de esta matriz se debe a que, para $N \neq 2$, todos los valores de la matriz son diferentes a cero. Esto significa que cada línea de retardo realimenta a otra línea de retardo, lo que ayuda a maximizar la densidad del eco (densidad temporal) rápidamente. La atenuación está dada por los filtros de absorción h_i (tratados en la siguiente sección).

La elección de los retardos m_i de cada línea es otro factor crucial. Típicamente se eligen números coprimos. Estos valores m_i deben asegurar una alta densidad de modos en todas las bandas de frecuencias. Si hay una densidad insuficiente, el sonido se puede oír como un “ringing” (coloración) o como una modulación irregular en amplitud (Tommasini, Ramos, y Ferreyra 2008a; Smith III 2010).

En de esta tesis, se sugiere una red con $N = 8$ líneas donde

$$\mathbf{m} = [601, 691, 773, 839, 919, 997, 1061, 1129]^T. \quad (4.17)$$

4.8.4 Ajuste a los tiempos de reverberación requeridos

Partiendo de un prototipo de red “sin pérdidas”, que tiene todos los polos sobre el círculo unitario en el plano z y su tiempo de reverberación es infinito, se busca ajustarlo a un tiempo de reverberación deseado. Es necesario mover los polos ligeramente hacia dentro del círculo unitario. Además, dicho tiempo también varía con la frecuencia. Por ejemplo, los polos en alta frecuencia están más amortiguados que los de baja frecuencia (Moorer 1979).

Jot y Chaigne (1991) indican que si la magnitud de los polos no es la misma, quedarán algunas frecuencias de resonancia. Para que cada línea tenga la misma relación de caída, la magnitud de los polos debe ser igual. Esto se logra haciendo que cada filtro h_i mantenga la siguiente relación

$$\gamma(\omega) = |h_i(e^{j\omega})|^{\frac{1}{m_i}} \quad \forall i. \quad (4.18)$$

Estos filtros, que se conectan luego de cada línea de retardo, son los que controlan el tiempo de decaimiento de la respuesta de la FDN. A partir de (4.18) se puede deducir la ecuación que cada filtro debe respetar

$$20 \log_{10} |h_i(e^{j\omega T})| = -60 \frac{m_i T}{T_{60}(\omega)} \quad (4.19)$$

entonces

$$|h_i(e^{j\omega T})| = 10^{\frac{-3m_i T}{T_{60}(\omega)}} \quad (4.20)$$

donde ω es la frecuencia normalizada, con $0 \leq \omega = 2\pi fT \leq 1$, f es la frecuencia en hertz, T es el período de muestreo y $T_{60}(\omega)$ especifica los tiempos de reverberación deseados versus la frecuencia. Estos tiempos fueron previamente estimados mediante la fórmula de Norris-Eyring (ver Capítulo 2).

Con filtros de primer orden de un polo (one-pole) de la forma

$$h_i(z) = \frac{g_i(1 - p_i)}{1 - p_i z^{-1}} \quad (4.21)$$

se obtienen buenos resultados. Tienen la ventaja que la componente continua es siempre $h_i(1)=g_i$ para todos los valores estables de p_i . Sin embargo, se puede hacer que g_i brinde un tiempo de reverberación deseado, mientras que p_i determina la razón de caída para las altas frecuencias. Las fórmulas de cálculo son:

$$g_i = 10^{\frac{-3m_i T}{T_{60}(0)}} \quad (4.22)$$

$$p_i = \frac{\ln(10)}{4} \log_{10}(g_i) \left(1 - \frac{1}{\alpha^2}\right) \quad (4.23)$$

siendo

$$\alpha \stackrel{\text{def}}{=} \frac{T_{60}(\pi/T)}{T_{60}(0)} \quad (4.24)$$

que denota la relación entre el tiempo de reverberación a la mitad de la frecuencia de muestreo dividido por el tiempo de reverberación en la corriente continua (Smith III 2010).

4.8.5 Corrección tonal

Un efecto adicional de los filtros absorbentes es la modificación de la energía de cada modo normal. Para compensar este efecto, Jot introduce un filtro corrector de

tono $t(z)$ que se coloca en serie con la FDN.

Para el caso de los filtros de primer orden usados, un buen corrector tonal está dado por el siguiente filtro de un cero (one-zero):

$$t(z) = \frac{1 - bz^{-1}}{1 - b} \quad (4.25)$$

donde

$$b = \frac{1 - \alpha}{1 + \alpha} \quad (4.26)$$

con α definida en (4.24).

4.8.6 Actualización de la FDN en el dominio del tiempo

Una vez calculados todos los parámetros necesarios para la FDN, se procede a la generación de la salida. Se utilizan las siguiente fórmulas para la actualización de los valores en cada paso de tiempo (ver Figura 4.18)

$$y[n] = t' \left[\sum_{i=1}^N c_i s_i[n - m_i] \right] + dx[n] \quad (4.27)$$

$$s_i[n] = \sum_{j=1}^N a_{ij} h'_i [x_j[n - m_i]] + b_i x[n] \quad (4.28)$$

donde $h'_i = \mathcal{F}^{-1} \{h_i\}$ y $t' = \mathcal{F}^{-1} \{t\}$, siendo \mathcal{F}^{-1} la transformada inversa de Fourier.

Las ganancias usadas son: $\mathbf{b} = [1, 1, \dots, 1, 1]^T$, $\mathbf{c} = [1, -1, \dots, 1, -1]^T$ y $d = 0$.

4.9 Síntesis de la RIR

Cuando la RIR es modelada con un método híbrido, se asume que se requiere un tiempo de transición específico entre las reflexiones tempranas y las reflexiones tardías. 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.

Este tiempo de transición debe ser tal que asegure que todas las características de las reflexiones tempranas permanezcan sin cambios (Lehmann y Johansson 2010). Según Stewart y Sandler (2007): “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”. Las reflexiones tempranas y la cola de reverberación deben ser empalmadas de manera que la transición sea continua e imperceptible al oyente.

Algunos autores sugieren diferentes valores (Jot 1992; Begault 1994; Garas 2000; Blesser 2001) y otros lo definen a partir de estudios psicoacústicos (Meesawat y Hammershoi 2002). En definitiva, este tiempo dependerá del recinto. Hay varios métodos que, a partir de una RIR conocida, permiten estimar el punto de transición (Stewart y Sandler 2007; Hidaka, Yamada, y Nakagawa 2007), pero el desafío es encontrar dicho punto sin conocer la RIR.

Basándose en algunos análisis estadísticos de la reverberación, Jot et al. (1997) establecieron que el tiempo a partir del cual la RIR se asemeja a un proceso estocástico es

$$t_{\text{mixing}} \approx \sqrt{V} \quad [\text{ms}] \quad (4.29)$$

conocido como *mixing time* (tiempo de mezclado), donde V el volumen del recinto en m^3 . En esta tesis se escogió dicho método para calcular el tiempo para unir ambos tipos de reflexiones.

A partir del valor de t_{mixing} , se calculó la distancia r que recorrerá un rayo sonoro en ese tiempo viajando a la velocidad del sonido c ($r = ct_{\text{mixing}}/1000$). Entonces, las reflexiones tempranas se computaron tomando como distancia

máxima a r . La amplitud inicial de las reflexiones tardías (de la envolvente) se estableció a $1/r$.

5 – Oyente

Las ondas producidas por una fuente sonora, luego de viajar por el medio de transmisión y reflejarse en los objetos del entorno, arriban al oyente. Antes de alcanzar los tímpanos, sufren alteraciones producidas por el torso, hombros y, especialmente, por la cabeza y el oído externo (pabellones de las orejas y conductos auditivos). Estas transformaciones son diferentes para cada oído y dependen del ángulo de incidencia de la onda. Dicha información permite localizar con precisión una fuente sonora en el espacio (Wightman y Kistler 1989a; Wightman y Kistler 1989b; Blauert 1997).

5.1 Localización auditiva

Como se ha discutido en el Capítulo 2, ha sido de interés para los investigadores en psicoacústica, estudiar la naturaleza de tales transformaciones y su implicancia en la audición espacial. Los primeros intentos se remontan al siglo pasado cuando Lord Rayleigh enunció la teoría Duplex. Firestone (1930) realizó un abordaje teórico-experimental considerando a la cabeza como una esfera rígida y analizó los niveles de presión sonora en dos polos separados por 180° (emulando la ubicación de los oídos). Concluyó que la diferencia de presión sonora entre ambos polos, debido a la difracción provocada por la cabeza, es función de la frecuencia y

de la dirección de la onda incidente.

La diferencia de presión sonora entre el oído izquierdo y el derecho se conoce como diferencia de nivel interaural (ILD). Las ILDs son despreciables por debajo de los 500 Hz, donde la longitud de onda es mayor al diámetro de la cabeza (Figura 5.1a), y crecen rápidamente hasta alcanzar su valor máximo alrededor de los 4000 Hz, cuando la fuente se encuentra enfrentada a uno de los oídos ($\pm 90^\circ$) (Figura 5.1b).

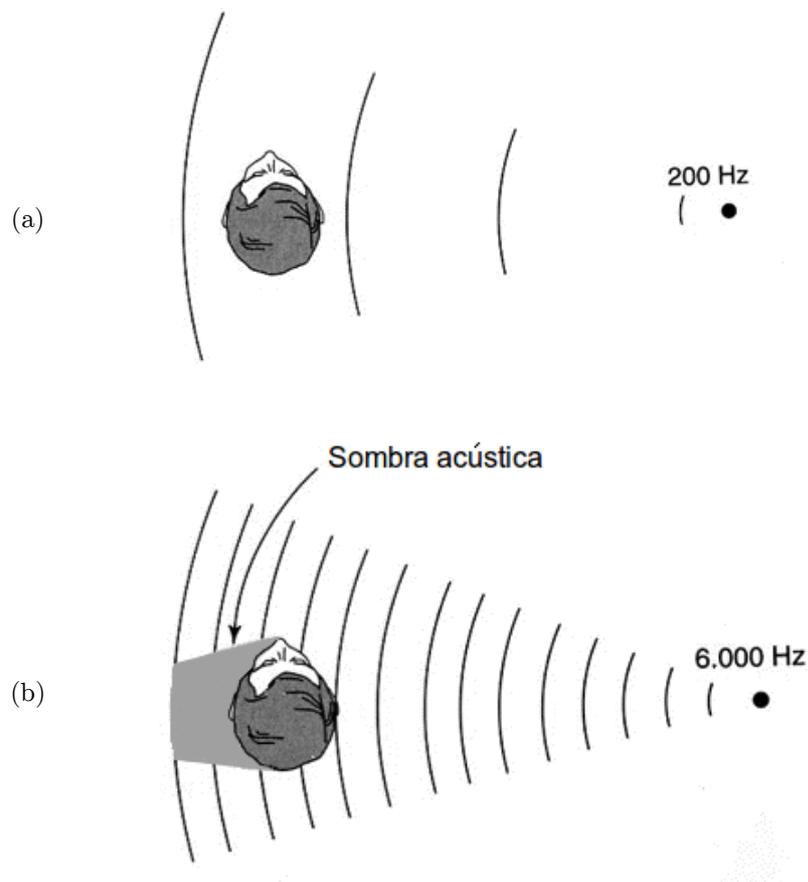


Figura 5.1 Ejemplo de la ILD (a) para baja frecuencia y (b) para alta frecuencia.

Adaptado de Goldstein (2005).

Continuando con el modelo esférico de la cabeza, Kuhn (1977) realizó estudios teóricos, que validó con mediciones, sobre la diferencia de los tiempos de arribo de la onda sonora a ambos oídos. Esta diferencia de tiempo se conoce como diferencia de tiempo interaural (ITD). Dedujo que las ITDs son dependientes del ángulo de incidencia y levemente de la frecuencia (Figura 5.2). Son mayores para bajas frecuencias (< 500 Hz); mínimas para frecuencias medias (entre 1000 y 1500 Hz) y aumentan para altas frecuencias (> 3000 Hz).

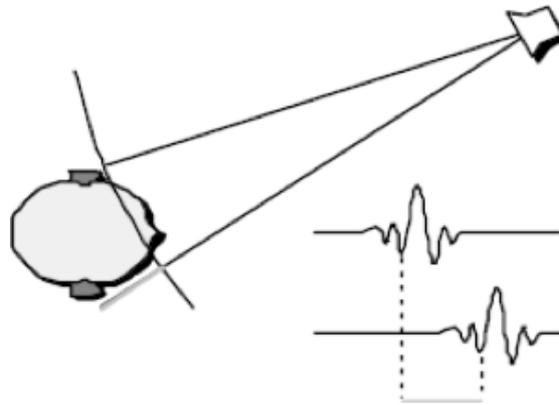


Figura 5.2 Ejemplo de la ITD

Tanto la ITD como la ILD, son los principales indicios para localizar una fuente sonora en el plano horizontal. Estas dos claves binaurales son conocidas como *claves primarias*.

En el plano medio, la eficacia de estas claves se ve reducida debido a que ambas son prácticamente cero. Una situación similar se presenta cuando la fuente sonora se ubica en un cono imaginario que se extiende hacia el exterior de cada oído a lo largo del eje interaural, que se conoce como *cono de confusión* (Wallach 1939) (Figura 5.3). Aunque la asimetría en la posición de las orejas en la cabeza y en las formas del pabellón auricular ofrecen algunas claves sonoras adicionales, no

es posible identificar las fuentes situadas en la superficie del cono de confusión y sólo se pueden resolver utilizando claves espectrales asociadas al filtrado natural que realiza el cuerpo humano.

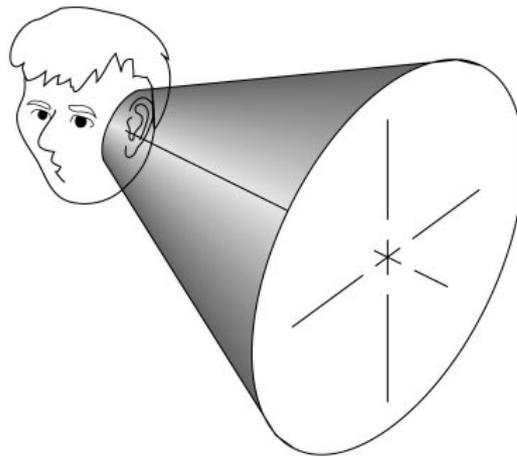


Figura 5.3 Cono de confusión

Debido a que las ondas sonoras se dispersan, refractan y reflejan en las protuberancias y cavidades de la orejas, originan una serie de picos y valles para frecuencias mayores a 4000 Hz, cuyas frecuencias centrales se desplazan sistemáticamente con el ángulo de incidencia de la onda sonora (Shaw 1974; Kuhn 1977).

Entonces, las orejas, los hombros y el torso juegan un rol preponderante, transformando el espectro de las ondas sonoras según la dirección de incidencia. Estas *claves secundarias* son denominadas claves monoaurales y son importantes para determinar la elevación de una fuente sonora.

Individuos diferentes producirán transformaciones diferentes a las claves

mencionadas. Por ejemplo, las ITDs y las ILDs varían con el diámetro de la cabeza, y los espectros con el tamaño y orientación de las orejas, entre otras razones.

En el contexto de un sistema de simulación acústica virtual es crucial modelar estas características. Más aún, si el sistema es dinámico e interactivo se deben asumir algunos compromisos para que el modelo sea eficiente y se ejecute en un tiempo aceptable.

La habilidad de los seres humanos para localizar fuentes sonoras en el espacio es óptima si la fuente se encuentra ubicada dentro de su campo visual. Sin embargo, en situaciones donde la fuente no es visible, los movimientos de cabeza tienen un rol esencial. Esta actitud, espontánea y natural, provoca variaciones continuas en el ángulo de incidencia de las ondas sonoras y ayuda a resolver ambigüedades en la posición de una fuente sonora (Perrott et al. 1990; Arias et al. 2006; Ramos, Arias, Tommasini, et al. 2009). Por lo expuesto, es fundamental sensor los movimientos de la cabeza para la simulación acústica virtual. Los movimientos de cabeza se pueden descomponer en tres tipos: rotación (izquierda-derecha), “tip” (arriba-abajo) y pivoteo (acercando la oreja hacia hombro del mismo lado) (Figura 5.4).

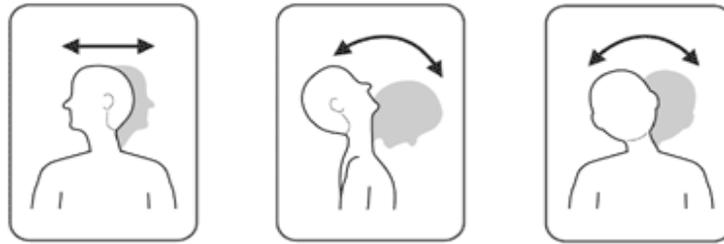


Figura 5.4 Tipos de movimientos de cabeza: rotación, tip y pivoteo

5.2 HRIRs y HRTFs

Las transformaciones de las ondas sonoras mencionadas, se encuentran presentes en las respuestas impulsivas de cabeza (HRIRs). Las HRIRs se obtienen mediante la medición en cámara anecoica (libre de reflexiones) de la respuesta impulsiva entre una fuente sonora y ambos tímpanos de un oyente. Se determinan para diferentes posiciones discretas de la fuente ubicada sobre una superficie esférica. En la proximidad de los tímpanos del sujeto (situado en el centro de una esfera) se colocan micrófonos miniaturas; como fuente, se emplean parlantes y como señales impulsos angostos, series de máxima longitud (Rife y Vanderkooy 1989) ó códigos de Golay (Zhou, Green, y Middlebrooks 1992).

Aplicando la transformada de Fourier a las HRIRs se obtiene las funciones de transferencia de cabeza (HRTFs), en el dominio de la frecuencia.

A modo de ejemplo, en la Figura 5.5, se muestran las respuestas impulsivas y las magnitudes espectrales de ambos oídos para la posición $(\theta, \phi) = (+45^\circ, 0^\circ)$ (de acuerdo al sistema de coordenadas referido al eje interaural)¹⁰.

¹⁰ Los sistemas de referencia de la cabeza están descriptos en el Apéndice B.

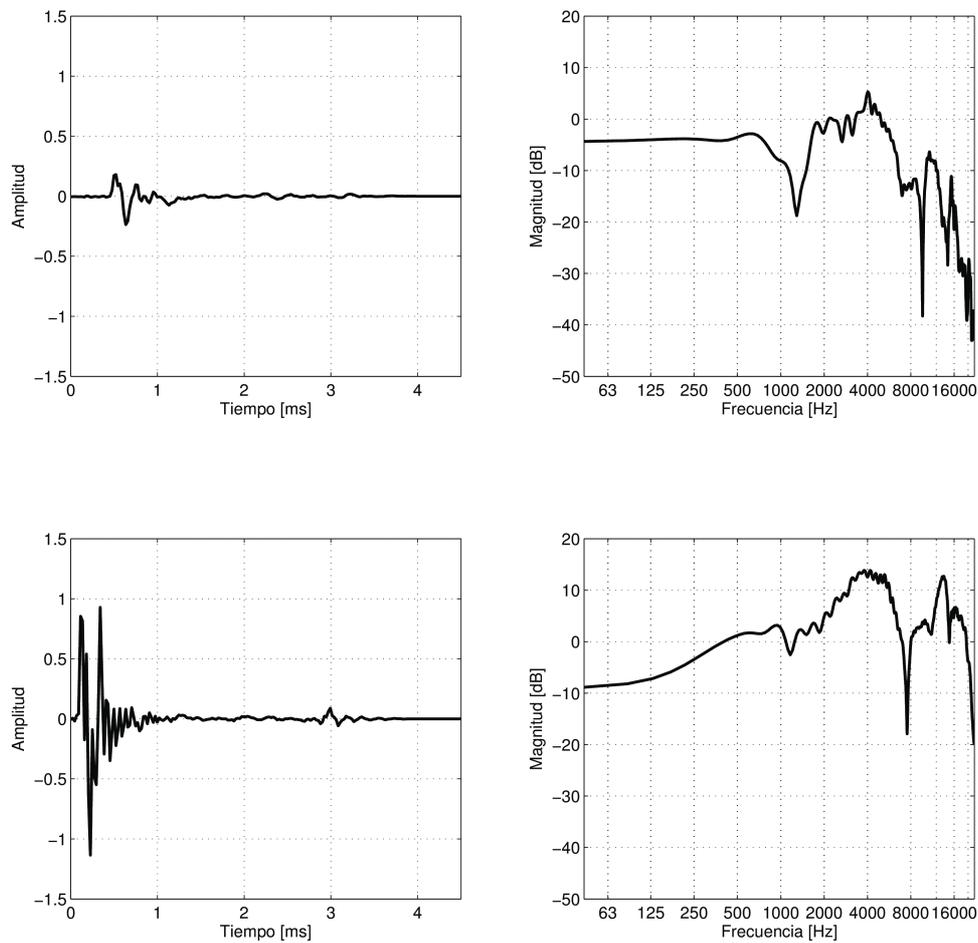


Figura 5.5 Respuesta impulsiva y magnitud de la función de transferencia del oído izquierdo (arriba) y del oído derecho (abajo) para una fuente sonora ubicada en azimut $+45^\circ$ y elevación 0° .

5.3 Influencia del oído externo en las HRTFs

La influencia del oído externo en las HRTFs se manifiesta para frecuencias cuyas longitudes de onda son comparables con sus dimensiones. El aporte que hace al espectro de las HRTFs se encuentra por encima de los 4 kHz. A continuación se discute la influencia en las HRTFs de las partes del oído externo.

5.3.1 Canal auditivo

Como se dijo más arriba, las mediciones de las HRIRs se suelen realizar con micrófonos miniaturas insertados en el conducto auditivo cercano a la membrana timpánica. Otros autores miden las HRIRs a la entrada del conducto auditivo bloqueado. Esta técnica evita los riesgos que implica introducir y ubicar un micrófono miniatura a pocos milímetros del tímpano. De todas maneras ambas mediciones son equivalentes ya que las modificaciones que produce el conducto auditivo en las ondas sonoras son independientes de la posición de la fuente (Shaw 1974).

5.3.2 Orejas

Shaw (1997) individualizó seis resonancias características, que se originan en las orejas (también denominada *pinna*) y las llamó modos normales. A partir de mediciones realizadas a 10 sujetos determinó valores medios de las frecuencias de resonancia, los niveles y las direcciones de excitación (Figura 5.6).

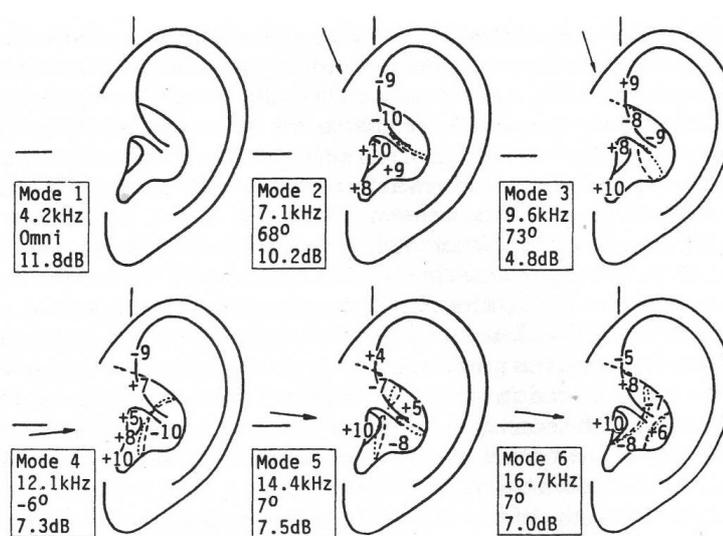


Figura 5.6 Modos normales identificados por Shaw

Según Shaw, el modo 1 es independiente de la dirección de excitación (omnidireccional) y se produce entre los 4 y 5 kHz, dependiendo del sujeto. Los modos 2 y 3, denominados modos verticales, se activan para elevaciones promedio de la fuente sonora de 68° y 73° , en las frecuencias centrales de 7,1 kHz y 9,6 kHz, respectivamente. Los denominados modos horizontales 4 y 5 se producen para elevaciones promedio de la fuente sonora de -6° y 7° a frecuencias de 12,1 kHz y 14,4 kHz respectivamente.

En un estudio previo (Shaw y Teranishi 1968), habían identificado las zonas de las orejas (Figura 5.7) donde se producen las resonancias mencionadas y desarrollaron un modelo mecánico aproximado utilizando formas geométricas simples (Teranishi y Shaw 1968). Posteriormente el modelo fue perfeccionado y fue denominado IRE (inclined rectangular cavity: stage E) (Shaw 1997). En este modelo, la concha fue representada por un cilindro con uno de sus extremos cerrados (el tímpano). Comprobaron que la frecuencia central del modo 1 (4,2 kHz), presente en las HRTFs en todas las posiciones, se correspondía con $\lambda/4$, siendo λ la profundidad del cilindro (22 mm). Cuando el cilindro fue reemplazado por una cavidad rectangular, el primer modo horizontal (modo 4 en 12,1 kHz) aparece y puede ser sintonizado variando la longitud transversal de la cavidad rectangular (≈ 17 mm). Para producir los modos 2 (7,1 kHz) y 5 (14,4 kHz), fue necesario introducir una barrera en la cavidad rectangular que tenga en cuenta la cruz helias.

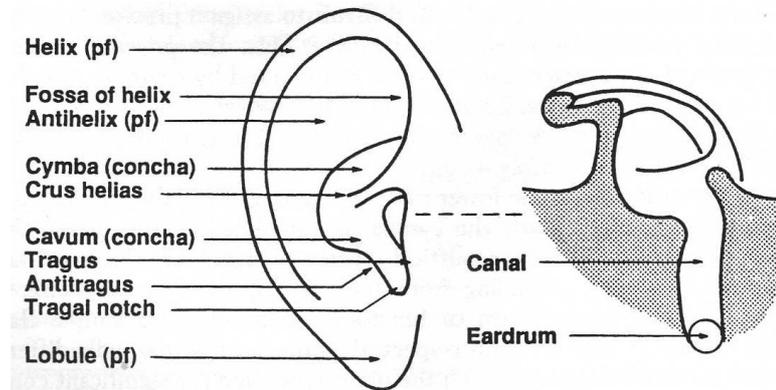


Figura 5.7 Zonas de la oreja. Extraído de Shaw (1997).

Otra característica de los espectros de las HRTFs atribuida a la oreja, es la presencia de mínimos notables (notchs) para frecuencias superiores a los 5 kHz. Se producen por la interacción de la onda directa que arriba a la entrada del conducto auditivo y por las reflexiones en diferentes lugares de la oreja. Son preponderantes para elevaciones entre -45° y $+45^\circ$ y desaparecen progresivamente cuando la fuente se eleva hasta $+90^\circ$.

El primer notch (N1) cuya frecuencia central es mayor a los 5 kHz se debe a reflexiones producidas en la pared posterior de la concha (Hebrank y Wright 1974; Asano, Suzuki, y Sone 1990; Lopez-Poveda y Meddis 1996; Iida et al. 2007). La frecuencia de este notch varía sistemáticamente con la posición (entre -45° y $+45^\circ$). Un segundo notch (N2) aparece entre -45° y 0° , su frecuencia es casi constante y se debe, probablemente, a reflexiones provenientes de la crus helias (Lopez-Poveda y Meddis 1996; Ramos, Tommasini, y Araneda 2010).

Para ilustrar estos fenómenos acústicos originados en las orejas, en la Figura 5.8 se muestran unas HRTFs procesadas. Las HRTFs corresponden al oído derecho de un sujeto. El eje vertical representa la frecuencia, el eje horizontal la posición de la fuente en el plano medio y los niveles de grises la presión sonora en decibeles. De

acuerdo con esta escala, las zonas más brillantes corresponden a niveles de mayor presión sonora. El procesamiento utilizado para extraer de las HRTFs sólo la contribución de las orejas se puede consultar en Ramos, Tommasini, y Araneda (2010). Explicado brevemente, consiste en aplicar el método estadístico de análisis de componentes principales (principal component analysis, PCA) a un grupo de HRTFs y reconstruirlas utilizando sólo las cuatro componentes principales de mayor varianza.

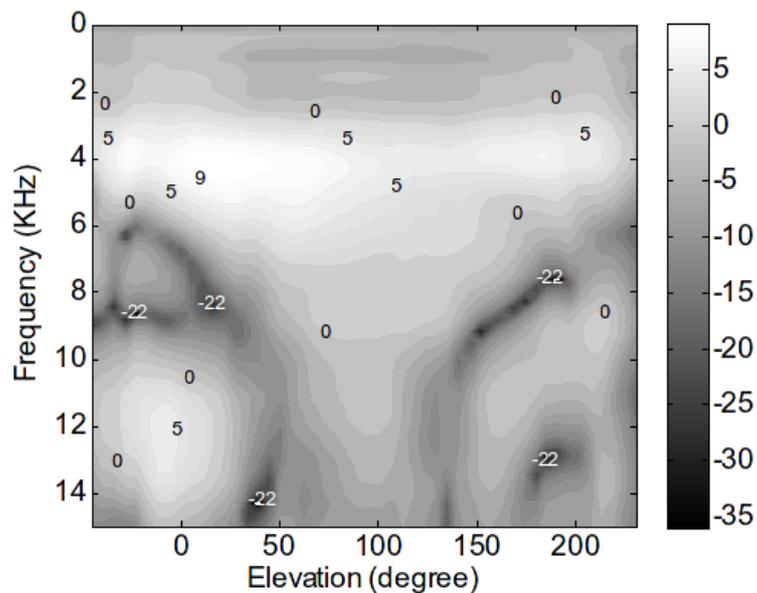


Figura 5.8 HRTFs para el plano medio. Extraído de Ramos, Tommasini, y Araneda (2010).

Como puede observarse, alrededor de los 4 kHz se aprecia una ancha resonancia que es independiente de la posición de la fuente, correspondiente al modo 1. Ésta se produce en la base de la cavum concha (Shaw 1997). La zona brillante entre los 11,5 kHz y 13,5 kHz a 0° aproximadamente (5 dB), es consecuencia de la

interacción de los modos horizontales 4 y 5 (12,1 kHz a -6° y 14,4 kHz a $+7^\circ$). Los modos 2 y 3 son difíciles de visualizar en la figura.

Entre los 6 kHz y los 10 kHz se observan dos mínimos (notchs N1 y N2) que varían sistemáticamente con la posición, entre -45° y $+45^\circ$. Como se anticipó, este mínimo se origina por difracción-reflexión en la pared posterior de la cavum concha (Shaw y Teranishi 1968; Lopez-Poveda y Meddis 1996). El desplazamiento sistemático de la frecuencia central de este mínimo, tiene un rol preponderante en la localización de fuentes sonoras en el plano medio (Blauert 1997; Iida et al. 2007). Entre $+150^\circ$ y $+180^\circ$ (detrás del sujeto) se observa un mínimo similar a los anteriores, simétrico respecto a $+90^\circ$ (arriba del sujeto) debido, probablemente, a la difracción producida en la parte posterior de la oreja.

5.4 Influencia de la cabeza en las HRTFs

Como se describió anteriormente, la cabeza es la responsable de generar dos claves primarias de localización: la ILD y la ITD. Dos siglos atrás, Lord Rayleigh enunció su modelo esférico de la cabeza (conocido como *Rayleigh's spherical model*, en inglés) al obtener la solución exacta de la difracción de una onda plana en una esfera rígida en dos puntos diametralmente opuestos (entrada a los conductos auditivos) (Brown y Duda 1998).

En la Figura 5.9 se muestra la variación del espectro para el oído izquierdo, en función del ángulo de incidencia y el radio de la cabeza. Es necesario aclarar que, para esta figura, ángulos positivos significan que la fuente está ubicada en el hemisferio del oído analizado (oído ipsilateral) y ángulos negativos lo contrario (oído contralateral). Para azimut $\theta = +90^\circ$, la presión sonora en el oído izquierdo (punto negro) es máxima debido a que la fuente sonora está directamente

enfrentada al mismo. La presión sonora para cuando la fuente está delante y detrás del oído disminuye sistemáticamente, siendo mínima en 0° y $+180^\circ$. Analizando el mismo oído, pero para cuando la fuente se encuentra en el lado opuesto (azimut negativo) el valor máximo se observa en azimut $\theta = -90^\circ$, esto es, cuando la fuente está directamente enfrentada al otro oído (punto blanco). En posiciones intermedias, por delante y detrás del oído derecho (-45° y -135°) la presión sonora disminuye y presenta una serie de ondulaciones (ripples) debido a que, según Shaw y Teranishi (1968) y Duda y Martens (1998), las ondas incidentes alcanzan al oído oculto (el derecho) por dos caminos: uno alrededor del frente y otro alrededor de la parte posterior de la cabeza. Los efectos de interferencia causados por la diferencias de estos caminos causan estos “ripples”.

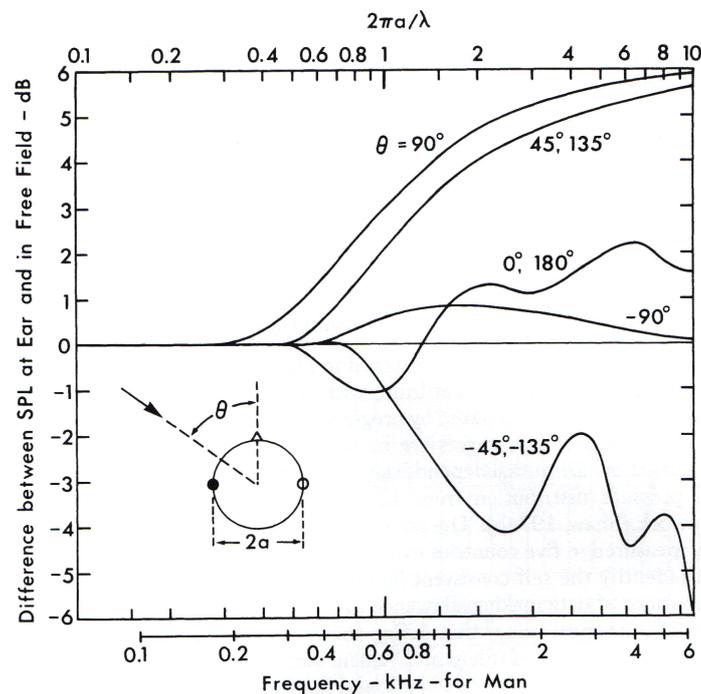


Figura 5.9 Variación del espectro para el oído izquierdo, en función del ángulo de incidencia y el radio de la cabeza.

Extraído de de Shaw (1997).

Para ilustrar estos efectos, en la Figura 5.10 se muestran las HRTFs del oído derecho del mismo sujeto anterior para el plano horizontal. La marcada diferencia entre los niveles de grises para azimuts positivos y negativos, revelan las diferencias de presión sonora del oído analizado (ILD). Cuando el oído está oculto por la cabeza (azimut negativo), se observan las ondulaciones mencionadas que se visualizan como máximos y mínimos, que se extienden desde las bajas frecuencias hacia las altas. y son simétricos respecto a -100° (la sombra acústica es mayor).

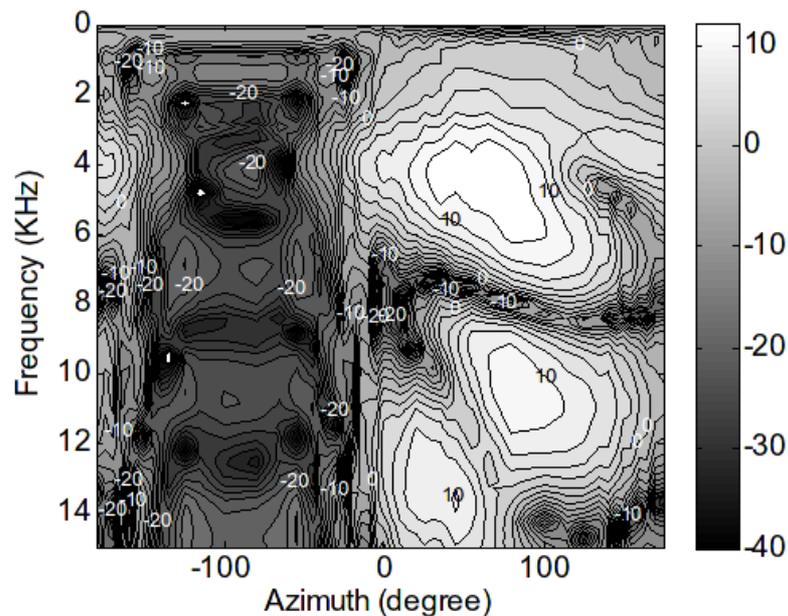


Figura 5.10 HRTFs para el plano horizontal. Extraído de Ramos, Tommasini, y Araneda (2010).

Para visualizar la ITD, en la Figura 5.11 se muestran las HRIRs (en el dominio del tiempo) para el oído derecho del mismo sujeto. El eje vertical representa el tiempo en milisegundos y el eje horizontal las posiciones de la fuente sonora en el plano horizontal. Para azimut positivos, las respuestas impulsivas presentan picos

de mayor energía (zonas más claras) que se producen al mismo instante de tiempo. Para azimut negativos (el oído derecho está oculto), los picos de mayor energía están retardados progresivamente, alcanzando 0,7 ms ($700 \mu\text{s}$) en -100° aproximadamente. En esta posición la ITD es máxima y está directamente relacionada con el diámetro de la cabeza del sujeto (ver Capítulo 2).

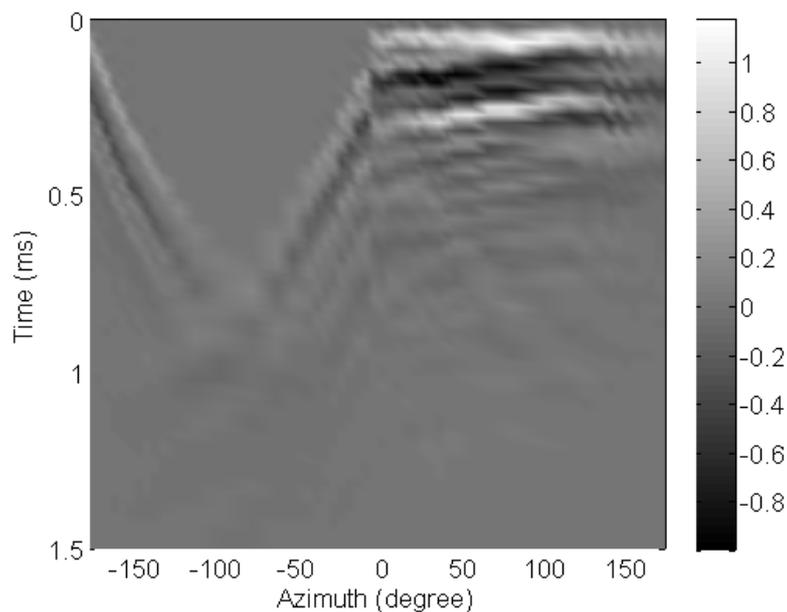


Figura 5.11 HRIRs para el plano horizontal. Extraído de Ramos, Tommasini, y Araneda (2010).

5.5 Influencia del torso y hombros en las HRTFs

Las reflexiones en torso y hombros se manifiestan en las bajas frecuencias, por debajo de los 4 kHz. Las ondas reflejadas arriban a las entradas de los conductos auditivos después de cierto tiempo, debido a los diferentes caminos que recorren. La interacción con la onda directa origina variaciones periódicas en el espectro de las HRTFs, típicas de un filtro peine (comb). Las reflexiones en el torso se

manifiestan en el plano medio, cuando la fuente se encuentra frente al sujeto y por debajo del eje interaural. Por el contrario, las reflexiones en los hombros son preponderantes cuando la fuente está situada encima del sujeto, esto es, elevación $+90^\circ$. Además, las ondas sonoras que se reflejan en el torso y hombros son modificadas por el oído externo y la cabeza antes de alcanzar los tímpanos de un sujeto (Ramos, Tommasini, y Araneda 2010).

5.6 Abordajes para el modelado del oyente

Desde finales de los 80 hasta nuestros días, la implementación eficaz de modelos del receptor humano acaparó el interés de numerosos investigadores. En general, se pueden distinguir tres tipos de abordajes:

- El físico-matemático teórico o estructural.
- El empírico.
- El funcional.

5.6.1 Abordaje estructural

Este abordaje, encabezado por Algazi et al. (2001; 2002), emplea modelos geométricos teóricos de la cabeza, torso y orejas, y modelan estos elementos con filtros paramétricos lineales. Han validado objetivamente los filtros implementados con mediciones de las ondas sonoras en cabezas y orejas artificiales.

5.6.2 Abordaje empírico

El abordaje empírico se basa en utilizar filtros digitales construidos a partir de HRIRs o HRTFs medidas. Si se considera a las HRTFs como las salidas de un sistema lineal e invariante en el tiempo (linear time-invariant, LTI), entonces pueden ser representadas por filtros. Aunque se debe notar que el problema no es

trivial.

La idea de modelar las HRTFs utilizando filtros digitales data del año 1989 (Wightman y Kistler 1989a; Wightman y Kistler 1989b). En estos estudios pioneros, los autores comprobaron que los sujetos podían localizar fuentes sonoras virtuales (sintetizadas mediante filtros FIR) usando auriculares con la misma precisión con que localizaban fuentes sonoras reales.

Luego, se comenzó a investigar la posibilidad de usar alternativas computacionalmente más eficientes, como son los filtros IIR (Sandvad y Hammershoi 1994; Huopaniemi y Karjalainen 1996; Blommer y Wakefield 1997; Haneda et al. 1999; Kulkarni y Colburn 2004; Ramos, Araneda, y Tommasini 2009a).

Mehrgardt y Mellert (1977) demostraron que las HRTFs pueden considerarse secuencias de fase-mínima y que el resto de fase es casi lineal según la posición e independiente de la frecuencia, y está asociado al ITD. Por lo tanto, las HRTFs pueden ser descompuestas en un sistema de fase-mínima y en un sistema pasa-todo (Oppenheim y Schaffer 1975). Esto permitió desarrollar un modelo simplificado conocido como: fase-mínima-más-retardo (minimum-phase-plus-delay). Algunos estudios psicoacústicos posteriores han corroborado que la habilidad de los sujetos para localizar una fuente sonora no se degrada si se utilizan HRTFs generadas por este modelo (Kistler y Wightman 1992; Kulkarni, Isabelle, y Colburn 1999).

Sin embargo, otros estudios más recientes demuestran que la fase de las HRTFs para bajas frecuencias (hasta 1500 Hz) cumple un rol dominante. Asumir que es lineal, sobre todo para el oído contralateral, conduce a errores insalvables en aplicaciones críticas (Scarpaci y Colburn 2005; Zahorik et al. 2006).

En un estudio reciente (Ramos, Araneda, y Tommasini 2009a), se modelaron

las HRIRs (en el dominio del tiempo) mediante filtros IIR de orden 20 por el método iterativo de Steiglitz-McBride (Steiglitz y McBride 1965), que tiene en cuenta tanto la magnitud como la fase (ver sección 5.8 más adelante).

5.6.3 Abordaje funcional

Finalmente, el tercer abordaje apela a métodos estadístico-matemáticos de reducción de datos. En lugar de trabajar con HRTFs individuales, se procesa un grupo extenso de HRTFs medidas a varios individuos con la finalidad de extraer aspectos comunes.

Es habitual el uso del análisis de componentes principales (PCA). El PCA transforma un número de variables correlacionadas en el mismo número de variables no correlacionadas, llamadas componentes principales (principal components, PCs) o funciones básicas. Además de reducir la dimensionalidad del conjunto de datos, resalta las características comunes y revela patrones ocultos (Shlens 2005).

Se obtiene un conjunto de PCs que, ponderadas por los coeficientes y combinadas linealmente, permiten reconstruir las HRTFs medidas para una posición y sujeto específico (Martens 1987; Kistler y Wightman 1992; Ramos, Calvo, y Tommasini 2007; Ramos y Tommasini 2008; Ramos, Araneda, y Tommasini 2008; Hwang y Park 2008).

5.7 Criterios de evaluación del modelo del oyente

Como el objetivo es obtener modelos eficientes que no degraden las claves perceptuales, es inevitable apelar a simplificaciones físico-matemáticas justificadas por estas evidencias. Los modelos que se obtienen son aproximaciones o estimaciones de las HRTFs reales.

En la literatura no hay acuerdo sobre la métrica adecuada para evaluar el error de ajuste, menos aún, sobre un error aceptable. Es habitual utilizar la distorsión espectral (spectral distortion, SD) para evaluar objetivamente los errores de la magnitud. La fórmula de cálculo es la siguiente (Nishino et al. 1999)

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(20 \log |H(f_i)| - 20 \log |\hat{H}(f_i)| \right)^2} \quad (5.1)$$

donde H es la HRTF medida, \hat{H} es la HRTF aproximada, y N es la cantidad de frecuencias discretas que cubren el ancho de banda de interés.

La SD puede ser un estimador adecuado para el plano horizontal, en donde la ITD y la ILD son relevantes. Pero, en el plano medio, no es confiable para evaluar la calidad de ajuste. Esto se debe a que, en este plano, los perfiles del espectro son muy relevantes, y este indicador oculta las diferencias espectrales por su carácter integrador respecto de la frecuencia.

Estrictamente hablando, la forma de validar los modelos es mediante pruebas psicoacústicas que comparen los rendimientos de personas estimuladas con señales auditivas reales y con señales virtuales generados por los modelos. Afortunadamente, en la actualidad, numerosos estudios psicoacústicos han logrado establecer, sin lugar a dudas, la implicancia perceptual de los detalles espectrales de las HRTFs. En el Capítulo 7, todas estas evidencias permitirán realizar un análisis objetivo de la calidad de ajuste de los modelos. Para evaluar objetivamente el modelo, se utilizó como criterio de aceptabilidad las mínimas diferencias notables (jnd) derivadas de estudios psicoacústicos mencionados anteriormente (por ejemplo, Moore, Oldfield, y Dooley 1989; Iida et al. 2007).

5.8 El método iterativo de Steiglitz-McBride

Esta tesis modela al oyente mediante filtros IIR utilizando el abordaje empírico. Con el método iterativo de Steiglitz-McBride es posible aproximar tanto la magnitud como la fase de las HRIRs dentro de valores perceptualmente aceptables.

Como se sabe, si a un sistema se le aplica una entrada

$$X(z) = \sum_{j=-\infty}^{+\infty} x[j]z^{-j}, \quad (5.2)$$

se obtendrá una respuesta

$$Y(z) = \sum_{j=-\infty}^{+\infty} y[j]z^{-j}. \quad (5.3)$$

El objetivo es diseñar un filtro IIR de la forma

$$H(z) = \frac{N(z)}{D(z)} = \frac{\sum_{k=1}^M b_k z^{-k}}{1 - \sum_{k=1}^N a_k z^{-k}}, \quad (5.4)$$

de manera que si se le aplica la entrada $X(z)$ se obtendrá una respuesta $W(z)$

$$W(z) = X(z) \frac{N(z)}{D(z)}. \quad (5.5)$$

Se debe minimizar el error cuadrático medio entre $W(z)$ y la respuesta original del sistema $Y(z)$ para encontrar los coeficientes a_k y b_k , de la forma

$$\sum E_j^2 = \frac{1}{2\pi j} \oint \left| X(z) \frac{N(z)}{D(z)} - Y(z) \right|^2 \frac{dz}{z}. \quad (5.6)$$

La minimización de esta ecuación implica un problema de regresión altamente no lineal. Steiglitz y McBride proponen una extensión a la técnica de Kalman donde tratan de minimizar la siguiente ecuación en sucesivas iteraciones

$$\oint \left| X(z) \frac{N_i(z)}{D_{i-1}(z)} - Y(z) \frac{D_i(z)}{D_{i-1}(z)} \right|^2 \frac{dz}{z}, \quad (5.7)$$

donde $i = 1, 2, 3, \dots$ es el número de iteraciones.

Se observa que si $D_{i-1}(z)$ converge a $D_i(z)$ la expresión (5.7) tiende a la (5.6). En cada iteración se filtran las señales $X(z)$ e $Y(z)$ con un filtro todo-polo con función de transferencia $1/D_{i-1}(z)$. En la primera iteración es necesario contar con los coeficientes a_k , los cuales se obtienen por cualquier método de diseño de filtros todo-polo.

En la práctica, el procedimiento descrito se realiza en el dominio del tiempo. En cada iteración se filtran las señales de entrada $x[j]$ y de salida $y[j]$ con el filtro todo-polo cuyos coeficientes a_k son calculados en la iteración anterior. Para obtener los valores iniciales de estos coeficientes se utiliza la recursión de Levinson (Levinson 1947).

Una vez obtenidas las señales filtradas $x_p[j]$ e $y_p[j]$ se define la siguiente ecuación lineal en diferencias

$$\begin{aligned} y_p[j] + a_1 y_p[j-1] + \dots + a_k y_p[j-k] = \\ x_p[j] + b_1 x_p[j-1] + \dots + b_k x_p[j-k] \end{aligned} \quad (5.8)$$

Si se despeja $y_p[j]$

$$\begin{aligned} y_p[j] = -a_1 y_p[j-1] - \dots - a_k y_p[j-k] + \\ x_p[j] + b_1 x_p[j-1] + \dots + b_k x_p[j-k] \end{aligned} \quad (5.9)$$

y se definen los vectores

$$\Theta = (a_1, \dots, a_k, 1, b_1, \dots, b_k)^T \quad (5.10)$$

$$\Phi[j] = (-y_p[j-1], \dots, -y_p[j-k], x_p[j], x_p[j-1], \dots, x_p[j-k])^T, \quad (5.11)$$

$y_p[j]$ queda

$$y_p[j] = \Phi^T[j] \Theta. \quad (5.12)$$

Esta ecuación lineal permite, en cada iteración, obtener el vector Θ , es decir, los coeficientes a_k y b_k a utilizar en la próxima iteración.

Como se verá en el Capítulo 7, con tan sólo 5 iteraciones se obtuvieron coeficientes que lograron muy buenos resultados.

5.9 Interpolación de HRTFs

En el mundo real, el movimiento relativo entre una fuente sonora y un oyente es percibido como un fenómeno auditivo continuo. Se ha demostrado que la gente adulta puede distinguir entre dos fuentes sonoras separadas entre 1° y 2° en el frente (Mills 1958; Blauert 1997). Pero, es prácticamente imposible medir las HRIRs o HRTFs con esa resolución espacial, debido principalmente al tiempo que se requeriría. Entonces, sólo se miden para algunos ángulos discretos de azimut y elevación.

Para poder recrear una verídica percepción continua del espacio circundante (donde es necesario cambiar de una HRTF a otra de acuerdo a los movimientos de la cabeza del oyente) se debe realizar algún tipo de interpolación. Sino al momento de reproducir un sonido continuo y realizar un cambio de HRTF, se escucharán “clics” audibles con ciertos estímulos (Tommasini y Ramos 2010).

Las primeras investigaciones sobre este tópico fueron realizadas por Wightman, Kistler y Arruda (1992), y Wenzel y Foster (1993), quienes estudiaron la influencia de la interpolación en la localización con HRTFs. Concluyeron que si no se interpolan las HRTFs pueden existir consecuencias perceptuales en el juzgamiento de la posición y confusiones delante-detrás y arriba-abajo.

Hay dos aspectos que deben ser discutidos: la resolución espacial mínima requerida para medir las HRTFs y la técnica de interpolación. Cualquier método

de interpolación incluirá errores, los cuales serán mayores si la resolución angular de las HRTFs medidas es pobre. Langendijk y Bronkhorst (2000) realizaron pruebas psicoacústicas para determinar la densidad y el método de interpolación más adecuado. El estudio se realizó para diferentes direcciones en el espacio (plano horizontal, plano medio, y posiciones diagonales). Compararon el rendimiento de los sujetos cuando utilizaban HRTFs interpoladas y HRTFs medidas en la misma posición. Determinaron que, aplicando interpolación bilineal a HRTFs medidas a una resolución espacial de alrededor de 6° , no se introducen claves audibles, y a una resolución entre 10° y 15° se introducen sólo algunas claves espectrales. Concluyeron que el rendimiento era el mismo, es decir, se localizaba la fuente sonora con igual precisión independientemente si la HRTF era medida o interpolada.

5.9.1 Interpolación bilineal

Aplicado este método en el dominio del tiempo, se obtiene la HRIR de una posición intermedia (no medida) mediante una combinación lineal de las cuatro HRIRs más cercanas, pesadas por coeficientes que dependen de la distancia de la posición interpolada a las medidas (Scarpaci 2006).

Se debe eliminar, por una parte, el retardo inicial (común a ambos oídos) dado por la distancia entre la fuente y los micrófonos al momento de la medición; y por la otra, el retardo dado por las ITDs. Este último retardo se debe guardar e introducir a posteriori. Luego, es necesario interpolar dos conjuntos de datos: las HRIRs y las ITDs.

Suponiendo que se quiere encontrar un valor desconocido de una HRIR para la posición $P = (\theta, \phi)$ para un tiempo t (Figura 5.12), y conociendo los valores para

las posiciones $Q_{11} = (\theta_1, \phi_1)$, $Q_{12} = (\theta_1, \phi_2)$, $Q_{21} = (\theta_2, \phi_1)$ y $Q_{22} = (\theta_2, \phi_2)$, se procede de la siguiente manera.

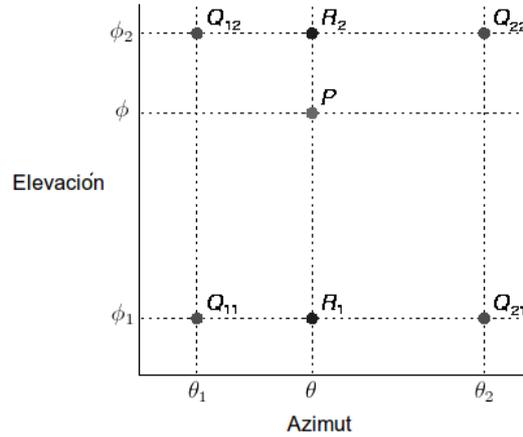


Figura 5.12 Puntos a calcular en una interpolación bilineal

Se realizan dos interpolaciones lineales en una misma dirección, para los puntos $R_1 = (\theta, \phi_1)$ y $R_2 = (\theta, \phi_2)$

$$\text{HRIR}(R_1, t) \approx \frac{\theta_2 - \theta}{\theta_2 - \theta_1} \text{HRIR}(Q_{11}, t) + \frac{\theta - \theta_1}{\theta_2 - \theta_1} \text{HRIR}(Q_{21}, t) \quad (5.13)$$

$$\text{HRIR}(R_2, t) \approx \frac{\theta_2 - \theta}{\theta_2 - \theta_1} \text{HRIR}(Q_{12}, t) + \frac{\theta - \theta_1}{\theta_2 - \theta_1} \text{HRIR}(Q_{22}, t). \quad (5.14)$$

Luego se procede a realizar otra interpolación lineal, en la otra dirección, para el punto P

$$\text{HRIR}(P, t) \approx \frac{\phi_2 - \phi}{\phi_2 - \phi_1} \text{HRIR}(R_1, t) + \frac{\phi - \phi_1}{\phi_2 - \phi_1} \text{HRIR}(R_2, t) \quad (5.15)$$

obteniéndose así el valor de la HRIR para el instante de tiempo t . Se debe proceder de la misma manera para el resto de las muestras y para el oído restante.

De una forma similar se interpolan los valores de las ITDs. En este caso, el

resultado de la interpolación es independiente del tiempo t y se realiza sólo una vez por cada par de HRIRs izquierda-derecha.

5.10 Personalización de HRTFs

Como se dijo anteriormente, cada HRTF contiene todas las transformaciones que se producen en una onda sonora antes de alcanzar los tímpanos de un oyente, al interactuar con la cabeza, la oreja, el torso y los hombros. Es diferente para cada oído y varía sistemáticamente con la ubicación de la fuente sonora. Cada HRTF es definida como un filtro acústico entre la fuente sonora y el tímpano del oyente (Blauert 1997). Se sabe, además, que no sólo depende de la ubicación de la fuente sonora, sino que también contiene información sobre el sujeto debido a las diferencias anatómicas tales como: tamaño de la cabeza y formas de la oreja; ancho de hombros y torso; entre otras. Estos indicios forman parte de la idiosincrasia del sujeto y su sistema auditivo los decodifica para localizar con precisión una fuente sonora en el espacio. Si las HRTFs utilizadas para modelar al oyente son las propias, la fuente se percibe compacta, externa y bien definida en una posición determinada del espacio. Por el contrario, si pertenecen a otro individuo, la fuente se percibe difusa, ubicada en el interior de la cabeza y aumentan las confusiones delante-detrás (Wenzel et al. 1993; Blauert 1997).

Para que un sujeto experimente una auténtica percepción del espacio, debe disponer de sus propias HRTFs. Sin embargo, estas mediciones son complejas, costosas y requieren de equipamiento especial, por ende es importante y necesario desarrollar procedimientos que permitan estimar las HRTFs de un individuo a partir HRTFs “genéricas” mediante otros abordajes más simples y menos costosos.

5.10.1 Primeros intentos de personalización de HRTFs

Históricamente, Kistler y Wightman (1992), utilizaron el PCA para aplicarlo a la magnitud de las HRTFs medidas a 10 sujetos (asumiéndolas como funciones de fase-mínima). Concluyeron, mediante pruebas psicoacústicas, que era posible modelar las HRTFs de cada sujeto, mediante la combinación lineal de cinco PCs. Las cuales tenían en cuenta el 90% de la varianza de las HRTFs originales de todos los sujetos. Un método similar empleó Chen et al. (1995), con la particularidad que lo aplicó tanto a la magnitud como a la fase.

Wenzel et al. (1993) utilizaron como modelo las HRTFs obtenidas de un sujeto que denominaron “buen localizador”, por su buen rendimiento en los estudios de Wightman y Kistler (1989a; 1989b). Concluyeron que los aciertos en el plano horizontal usando auriculares, eran comparables con los de campo libre. Sin embargo, eran menos exactos cuando evaluaban la elevación de una fuente ya que las confusiones delante-detrás y arriba-abajo aumentaban.

Middlebrooks (1999) llevó a cabo un abordaje interesante basándose en dos postulados: a) las diferencias en los perfiles espectrales (picos y notchs) de las HRTFs entre sujetos varían sistemáticamente con el tamaño de las medidas antropométricas del sujeto (cabeza, hombros y pabellones auriculares); y b) un factor de escala óptimo puede deducirse entre las HRTFs de dos sujetos a partir del diámetro de la cabeza y el tamaño de la cavidad del pabellón auricular. En otras palabras, un sujeto puede utilizar las HRTFs de otro sujeto, seleccionando de un banco de HRTFs aquellas cuyo factor de escala sea más pequeño. Se comprobó que los sujetos cometían menos errores cuando utilizaban HRTFs medidas a otros individuos seleccionadas por este método.

5.10.2 Métodos actuales de personalización de HRTFs

Actualmente, diferentes estudios han abordado el problema de la personalización de las HRTFs, desde diferentes enfoques. Para una reseña sobre los métodos se puede consultar el trabajo de Xu et al. (2007), que identificaron siete métodos potenciales de individualización. Generalmente estos métodos se basan en utilizar aquellos parámetros antropométricos del individuo que intervienen en la conformación de las HRTFs, por ejemplo, (Rodríguez y Ramírez 2005; Xu, Li, y Salvendy 2007; Hwang y Park 2007; Xu, Li, y Salvendy 2009; Hwang, Park, y Park 2008).

Una de las formas más utilizadas para introducir dichos parámetros en la construcción de HRTFs individuales es, en primer lugar, aplicar el método de PCA sobre la base de datos de HRTFs para obtener un conjunto de PCs llamadas genéricas.

En segundo lugar, se estiman (personalizan) los pesos del PCA para un sujeto a partir de una regresión lineal múltiple con las medidas antropométricas del grupo de sujetos (Rodríguez y Ramírez 2005; Hu et al. 2006; Ramos, Araneda, y Tommasini 2009b). Basándose en el hecho de que la influencia de la cabeza, pinna y torso es de característica compleja, otros autores apelan a algún método de regresión no lineal múltiple para estimar dichos pesos (Hu et al. 2008; Huang y Fang 2009).

Finalmente, a partir de los pesos estimados y las PCs genéricas, se reconstruyen las HRTFs personalizadas.

No está claro aún cuáles son las medidas de los rasgos anatómicos más dominantes en las características de las HRTFs, ni tampoco el método de selección. En general, es habitual determinar el grado de correlación que existe entre un gran

número de mediciones anatómicas y las HRTFs. Ahora, estas mediciones son unidimensionales y no tienen en cuenta las formas, como por ejemplo, la pared posterior de la concha que tiene un rol decisivo en la generación de los notchs, claves dominantes en el plano medio (Hebrank y Wright 1974; Lopez-Poveda y Meddis 1996; Iida et al. 2007). Para evaluar objetivamente la calidad del ajuste entre las HRTFs originales y las estimadas, es común utilizar la distorsión espectral (SD). Como se dijo anteriormente, la SD se define como la distancia euclidiana entre dos espectros integrada en todo el rango de frecuencias. Pero este índice tiene la particularidad de englobar y ocultar las diferencias de los perfiles espectrales, un aspecto clave para localizar fuentes sonoras en el plano medio.

También se realizaron avances en personalización de HRTFs. El trabajo tuvo como propósito construir HRTFs personalizadas a partir de la estimación de los pesos por métodos de regresión lineal y no-lineal multivariable y comparar los resultados en el plano medio. Se utilizó el PCA aplicado a un conjunto de HRTFs medidas a 30 sujetos. Mediante los métodos de regresión lineal múltiple y redes neuronales artificiales (artificial neural networks, ANNs), se estimaron los pesos del PCA para cinco sujetos externos al grupo previo¹¹. Para comparar los resultados, se reconstruyeron las HRTFs para el plano medio con seis PCs (que tenían en cuenta el 90% de la varianza) a partir de los pesos estimados por ambos métodos.

Del estudio se concluyó que, si bien los resultados indican que las ANNs son mejores que la regresión lineal múltiple, se mostró que la SD no es un índice apropiado para evaluar la calidad de ajuste entre las HRTFs medidas y las personalizadas en el plano medio. Los valores de los errores entre las HRTFs medidas y las HRTFs estimadas son detectables y discriminables y podrían

11 A diferencia de los artículos de referencia citados, como por ejemplo, (Hu et al. 2008).

producir consecuencias perceptuales severas. La utilización de medidas unidimensionales de las cavidades y protuberancias de la pinna, puede ser adecuado para estimar algunas de las resonancias características del oído externo. Sin embargo, para determinar las frecuencias centrales y presión sonora de los notchs característicos entre -45° y $+45^\circ$ ¹², es esencial que se reporte las formas de la pinna, en especial, la de la pared posterior de la concha.

¹² Rango donde los notchs son importantes.

6 – Implementación del sistema

En el presente capítulo se detalla cómo fue implementado el sistema de simulación acústica virtual en tiempo real (AVRS). El mismo proporciona una experiencia de sonido binaural mediante auriculares, permitiendo que el sujeto explore la escena mediante movimientos de cabeza, logrando la inmersión en un entorno acústico virtual.

6.1 Reseña sobre sistemas de simulación acústica

Existen varios sistemas de auralización disponibles comercialmente que fueron desarrollados como un componente adicional a programas que modelan la acústica de recintos. Ejemplos de estos sistemas son: EARS (Ahnert y Feistel 1992), ODEON (Naylor 1993), RAMSETE (Farina 1995), y CATT-Acoustic (Dalenbäck 1996).

También se encuentran softwares académicos no comerciales: el proyecto DIVA (Savioja et al. 1999) de TKK, Finlandia, que fue diseñado como un sistema de AVR, y luego fue unificado con el proyecto EVE de realidad virtual de la misma universidad; IKA-SIM (Silzle, Strauss, y Novo 2004) de la Universidad Ruhr de

Bochum es utilizado en investigaciones psicoacústicas, con el objetivo de convertirse en un sistema de base (backend) para diferentes aplicaciones Web; VirKopf/RAVEN (Lentz et al. 2007) de la Universidad RTWH de Aachen, combina la simulación de recintos con la simulación gráfica; REVES (Moeck et al. 2007) desarrollado en el INRIA, tiene la meta de calcular eficientemente la difusión y sintetizar una gran cantidad de fuentes sonoras y reflexiones simultáneamente (incluyendo gráficos); y el reciente proyecto AURALIAS (Bos y Embrechts 2009) de la Universidad de Liege de simulación acústica.

Todos estos sistemas son de código cerrado, lo que imposibilita profundizar en su funcionamiento y utilizarlos como base para nuevos estudios. Son pocos los proyectos de código abierto (open-source) que tienen como objetivo la auralización dinámica. Uno de ellos es slab3d (conocido anteriormente como SLAB) desarrollado inicialmente en la NASA (Wenzel, Miller, y Abel 2000). Brinda una espacialización 3D pero el modelado acústico de recintos está limitado solamente a recintos cúbicos. RTVAS (Scarpaci 2006) de la Universidad de Boston es el único de los proyectos open-source diseñado para tener en cuenta explícitamente las restricciones de tiempo impuestas por este tipo de sistemas. Se ejecuta sobre un *kernel* de tiempo real y su objetivo es estudiar la localización de sonidos, pero no cuenta con un módulo para modelado de recintos. Otro proyecto es Uni-Verse (Kajastila et al. 2007), que fue desarrollado por un consorcio de siete miembros europeos con el objetivo de crear una plataforma para Internet de gráficos y audio tridimensional. EVERTims (Noisternig et al. 2008) también de TKK, utiliza algunos algoritmos más avanzados que Uni-Verse. Por último, el software The SoundScape Renderer (SSR) (Ahrens, Geier, y Spors 2008) de la Universidad Técnica de Berlín es un *framework* para síntesis de audio espacial. Su arquitectura

modular permite el uso de diferentes métodos de síntesis del sonido auralizado.

Como se puede apreciar, estos sistemas tienen objetivos disímiles, que van desde el diseño de salas de concierto hasta la aplicación de pruebas específicas de localización. No es posible utilizarlos como base para propósitos diferentes para los cuales fueron diseñados, ni tampoco adaptarlos a necesidades específicas. Además, muchos de los proyectos de código cerrado no están disponibles libremente para su uso.

Son pocos los proyectos de código abierto (open-source) que tienen como objetivo la auralización dinámica e interactiva. El único que incluye el modelado acústico de recintos es EVERTimes. En general, trabajan sobre plataformas no específicas para la ejecución en tiempo real, por lo que es imposible controlar con precisión los tiempos de cálculo y las latencias. Solamente RTVAS es un proyecto de código abierto que trabaja sobre un sistema en tiempo real. De hecho, el sistema aquí propuesto tiene una estructura de alto nivel similar al RTVAS, aunque los modelos utilizados y la implementación son diferentes.

Actualmente, algunos de estos proyectos académicos no continúan activamente su desarrollo, ya que no se conocen publicaciones referidas a ellos en los últimos años; otros directamente fueron abandonados. Los proyectos que se encuentran activos han comenzado su fase de desarrollo hace unos cinco ó seis años aproximadamente, en paralelo con el sistema de simulación acústica virtual en tiempo real aquí presentado.

6.2 Sistemas en tiempo real

La realidad acústica virtual requiere de computación en tiempo real. Estos sistemas deben responder a eventos del entorno dentro de restricciones precisas de

tiempo. Como consecuencia, su correcto comportamiento no sólo depende del rigor de los modelos sino también del momento en que los resultados son producidos. Esto no significa que la respuesta deba ser rápida, sino simplemente que sea lo “suficientemente rápida” en el contexto en el cual opera el sistema (Buttazzo 2004). Por eso, la diferencia principal entre una tarea de tiempo real y una que no es de tiempo real, es que la primera posee un tiempo límite o máximo (deadline) en el cual debe completar su ejecución.

Para el caso del AVRS, el término “suficientemente rápido” está relacionado con el tiempo de latencia entre una acción del oyente y la actualización del escenario acústico. Como se verá en el Capítulo 7, esto depende de varios factores. En general, se puede tomar como cota máxima a una latencia que ronda los 92 ms para reflejar los cambios del escenario acústico (Wenzel 1998).

6.2.1 Tareas de tiempo real

Dependiendo de las consecuencias que acarrea el incumplimiento del deadline, las tareas de tiempo real se suelen clasificar en *hard real-time* (HRT) y *soft real-time* (SRT). Una tarea HRT tiene un cumplimiento estricto del deadline. Si lo supera podría causar serios problemas al entorno donde está operando el sistema. Para el caso del AVRS escuchar “clics” o discontinuidades en la reproducción del sonido. En cambio, en una tarea SRT no es necesario el cumplimiento estricto del deadline, pero sí es deseable por cuestiones de rendimiento. Si el deadline es superado, no causaría problemas graves en el entorno ni tampoco corre riesgo el correcto funcionamiento del sistema. Por ejemplo, el sensado de los movimientos de cabeza.

Un sistema operativo de tiempo real (real-time operating system, RTOS) debe garantizar que se cumplan todas estas restricciones críticas de tiempo. A menudo

se les suele llamar sistemas operativos de *hard real-time* a aquellos que soportan tareas HRT. Muchas aplicaciones, como los sistemas de realidad acústica virtual, incluyen tareas HRT y SRT. En estos casos, los RTOSs tienen por objetivo garantizar las restricciones de tiempo individuales de las tareas HRT, al mismo tiempo que minimizan los tiempos promedio de respuestas para las tareas SRT.

6.2.2 Planificación de las tareas de tiempo real

Un RTOS debe garantizar que se cumplan todas las restricciones críticas de tiempo. Una característica importante que deben poseer es la multitarea apropiativa (mejor conocida como *preemption*), que significa que el RTOS permite que los procesos que se ejecutan con máxima prioridad (tareas HRT) puedan interrumpir a cualquier otro proceso de menor prioridad (otras tareas) para acceder a los recursos que necesiten. Un planificador *preemptive* hace que los tiempos de respuesta sean predecibles. Esto es la principal característica del tiempo real: ser capaz de garantizar el tiempo máximo que se tardará en responder a un evento (Abbott 2006).

El principal componente que afecta la planificación (scheduling) es el procesador. Las características internas de éste, tales como los mecanismos de precarga de instrucciones (instruction prefetch), segmentación (pipelining), memoria caché y acceso directo a memoria (direct memory access, DMA), mejoran el rendimiento promedio del procesador pero introducen un factor no determinístico que hace difícil realizar análisis sobre el tiempo de ejecución del peor caso (Buttazzo 2004).

Otro factor influyente es la construcción interna del kernel de tiempo real, como el algoritmo de planificación en sí, los mecanismos de sincronización, los tipos de semáforos, la administración de memoria, la semántica de comunicación y el

mecanismo de manejo de las interrupciones.

6.2.3 GNU/Linux como RTOS

En los últimos años, el crecimiento de GNU/Linux (y del software libre en general) hizo que se convierta casi en un estándar de facto para las ciencias computacionales. Debido a que su kernel es libre, generó mucho interés la posibilidad de convertir este sistema operativo de propósito general en un RTOS.

Básicamente, existen dos abordajes para lograrlo. El primero implica aplicar un “parche” al kernel de GNU/Linux para hacerlo más preemptable¹³, lograr menores latencias en las interrupciones y en la planificación, mejorar los tiempos de los cambios de contexto (context switching) y tener mayor precisión en el temporizador. El segundo abordaje implica usar un nano-kernel de tiempo real con un planificador totalmente preemptable que ejecuta el kernel de GNU/Linux como una tarea de baja prioridad. Este nano-kernel también administra las interrupciones del hardware subyacente y sólo redirecciona hacia el kernel GNU/Linux las que son relevantes para el mismo (posee un *interrupt dispatcher*) (Vun, Hor, y Chao 2008).

Si bien existen ventajas y desventajas en cada uno de estos abordajes, el más aceptado es el del nano-kernel. Permite mayor control de las latencias y, por lo tanto, del determinismo. Dentro de este abordaje existen dos extensiones open source: RTAI¹⁴ (Real-Time Application Interface) (Mantegazza, Dozio, y

¹³ Si bien GNU/Linux (como también otros sistemas operativos) ya es preemptable, no se lo considera de tiempo real debido a que hay períodos excesivamente largos durante los cuales esta planificación no se aplica estrictamente.

¹⁴ www.rtai.org

Papacharalambous 2000) y Xenomai¹⁵.

Tanto RTAI como Xenomai están basados en el nano-kernel Adeos¹⁶ (Adaptive Domain Environment for Operating Systems), que agrega una capa de abstracción de hardware (hardware abstraction layer, HAL). Adeos se sitúa bajo el sistema operativo GNU/Linux en ejecución y luego provee los mecanismos necesarios para permitir que múltiples sistemas operativos (llamados dominios) compartan el mismo entorno de hardware (Yaghmour 2001). Uno de estos dominios es GNU/Linux y al cargarse los módulos de RTAI se registra un nuevo dominio (Bucher y Balemi 2005).

RTAI administra la planificación de las tareas de tiempo real y deja a GNU/Linux proveer el resto de la funcionalidad. Pero las interrupciones de hardware son interceptadas por RTAI. Así, esta extensión propaga, mediante el nano-kernel Adeos hacia el kernel GNU/Linux, solamente las interrupciones que no son de interés para las tareas de tiempo real (Figura 6.1). Estudios comparativos (Barbalace et al. 2008) indican que esta es una de las características que hace que RTAI logre un mejor rendimiento (menor *delay* y menor *jitter*) que Xenomai, e inclusive que otras alternativas comerciales no libres. La latencia del RTOS con RTAI es menor a 10 μs para una computadora estándar, pero obviamente este valor cambiará de acuerdo al hardware disponible.

15 www.xenomai.org

16 home.gna.org/adeos

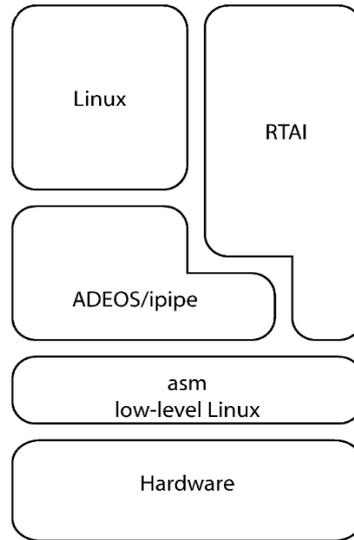


Figura 6.1: Estructura de capas de RTAI. Extraído de Barbalace et al. (2008).

LXRT (LinuX Real-Time) es un módulo de RTAI que extiende su funcionalidad al espacio de usuario¹⁷, lo que permite ejecutar tareas HRT en este espacio usando la misma API (application programming interface) provista por RTAI para el espacio de kernel. Esto lo logra gracias al planificador de RTAI y al bloqueo de interrupciones de hardware mientras una tarea HRT se encuentra en ejecución.

Las ventajas son significativas: permite definir en el mismo programa algunas tareas que se ejecutan en tiempo real y otras que no se ejecutan en tiempo real; y permite comunicar a las tareas de tiempo real con procesos de GNU/Linux estándares.

Para el correcto funcionamiento de los sistemas en tiempo real se deben evitar

¹⁷ El espacio de usuario es el área de memoria donde se ejecutan las aplicaciones del usuario. A diferencia del espacio de kernel, que es el área de memoria reservado para ejecutar el kernel y los módulos del kernel que hacen al funcionamiento interno de un sistema operativo.

algunas condiciones conocidas como “asesinas de la latencia” (latency killers). Los desarrolladores de RTAI indican varias (Racciu y Mantegazza 2006):

- actividades de acceso directo a memoria (direct memory access, DMA) demandantes (como el disco duro o algunos dispositivos PCI),
- uso intensivo de interfaz gráfica,
- soporte de versiones antiguas de USB,
- la administración de energía (tanto APM como ACPI),
- escalado de la frecuencia de la CPU,
- la administración de señales de interrupción (system management interrupt, SMI).

Estas condiciones causan resultados impredecibles en los tiempos, lo cual es incompatible con el concepto de tiempo real.

6.3 Desarrollo del AVRS

En primera instancia y con el objetivo de verificar los modelos, se ha desarrollado una serie de programas basados en *scripts* de MATLAB que se ejecutaban “off-line”. Los principales fueron:

a) Un simulador de acústica de recintos, con el cual es posible calcular la RIR y visualizar la trayectoria de los rayos calculados por el ISM (Tommasini y Ramos 2008) (ver Figura 6.2 y Figura 6.3).

b) Un módulo para calcular los parámetros acústicos especificados en la norma ISO 3382 (Tommasini, Ramos, y Ferreyra 2008b) que permitió evaluar las RIRs simuladas¹⁸ (Figura 6.4). Esto es debido a que, según el programa que se use para calcularlos, los valores de los parámetros varían, en algunos casos, de manera

¹⁸ Este módulo está basado en algunos algoritmos del software open-source AcMus (Iazzetta et al. 2004; Gomes de Queiroz et al. 2008).

alarmante. Las diferencias están relacionadas con el criterio que se adopte para determinar el nivel inicial, el límite de integración y el método utilizado para calcular la relación señal-ruido (signal-noise ratio, SNR). La ausencia de ruido de fondo en la RIR simulada puede ser la causa de algunas discrepancias.

c) Un software que incorporó varios modelos del oyente con diversos niveles de complejidad y permitió la administración de pruebas psicoacústicas para evaluarlos (Ramos, Calvo, y Tommasini 2007; Ramos, Araneda, y Tommasini 2009a) (Figura 6.5).

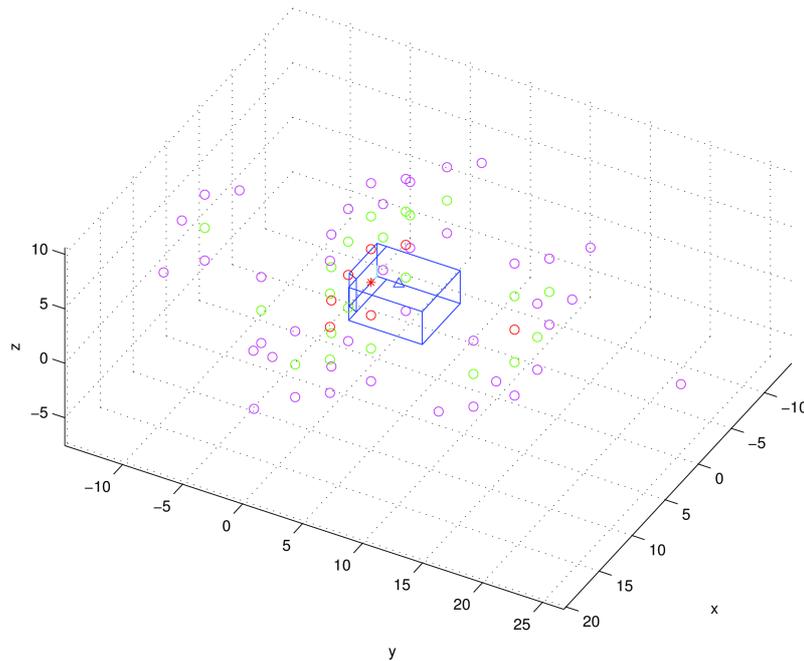


Figura 6.2 Cálculo y visualización de las VSs de un recinto.

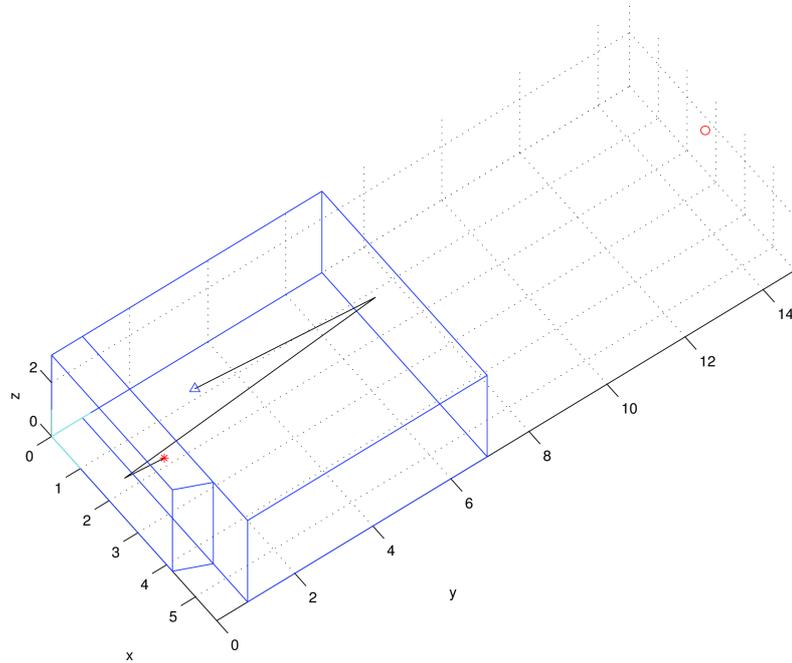


Figura 6.3 Visualización de la trayectoria de un rayo para una VS de segundos orden.

Los modelos, una vez validados, fueron migrados al sistema en tiempo real. Debido al alto costo computacional que implica realizar todos los cálculos de los modelos involucrados, el AVRS se ejecuta en una máquina estándar pero dedicada para tal fin, llamada servidor de auralización. Está implementado sobre un RTOS GNU/Linux usando RTAI-LXRT. Es un sistema basado totalmente en software y escrito en C/C++, bajo el paradigma orientado a objetos. Para evitar las condiciones llamadas “asesinas de latencia” se han desactivado dichas opciones tanto en la BIOS (basic input/output system) del servidor de auralización como en la configuración del kernel modificado.

No tiene ningún tipo de restricción con respecto a los requerimientos de hardware, como por ejemplo, la utilización de hardware DSP específico para

procesamiento de audio. Sólo es necesario un sensor de movimientos de cabeza (tracker) para proveer la realimentación de los movimientos de cabeza del usuario. A diferencia de otros sistemas de auralización (Savioja et al. 1999; Lentz et al. 2007), no necesita un montaje especial de una gran cantidad de altavoces. Sólo con unos auriculares es suficiente para su uso, haciéndolo mucho más portable.

Como el sistema es open-source y el hardware en el que se ejecuta es de un costo relativamente bajo, abre la posibilidad de que otros laboratorios estudien, implementen, modifiquen y expandan el sistema con un mínimo esfuerzo.

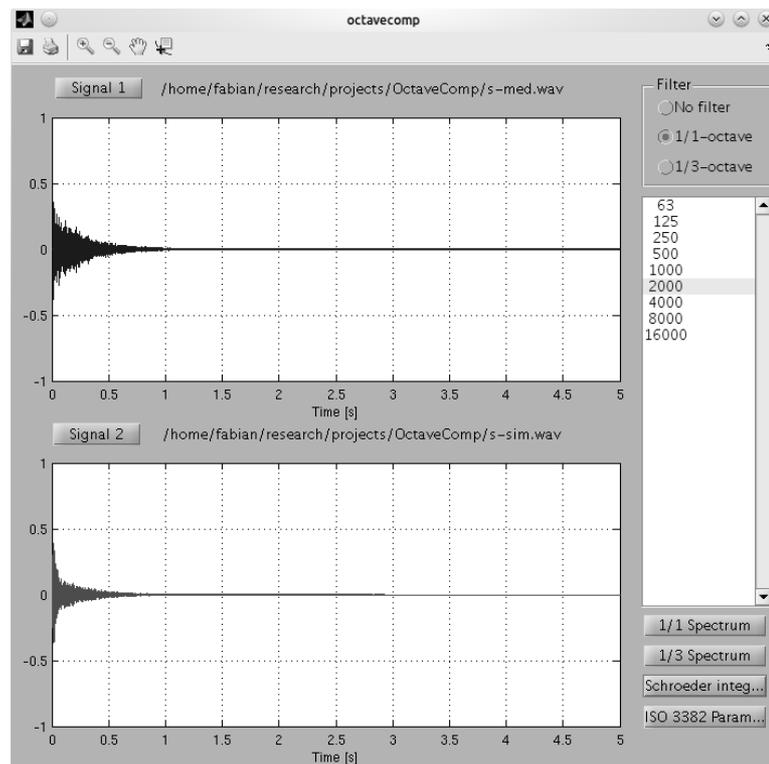


Figura 6.4 Comparación objetiva de RIRs medidas y simuladas

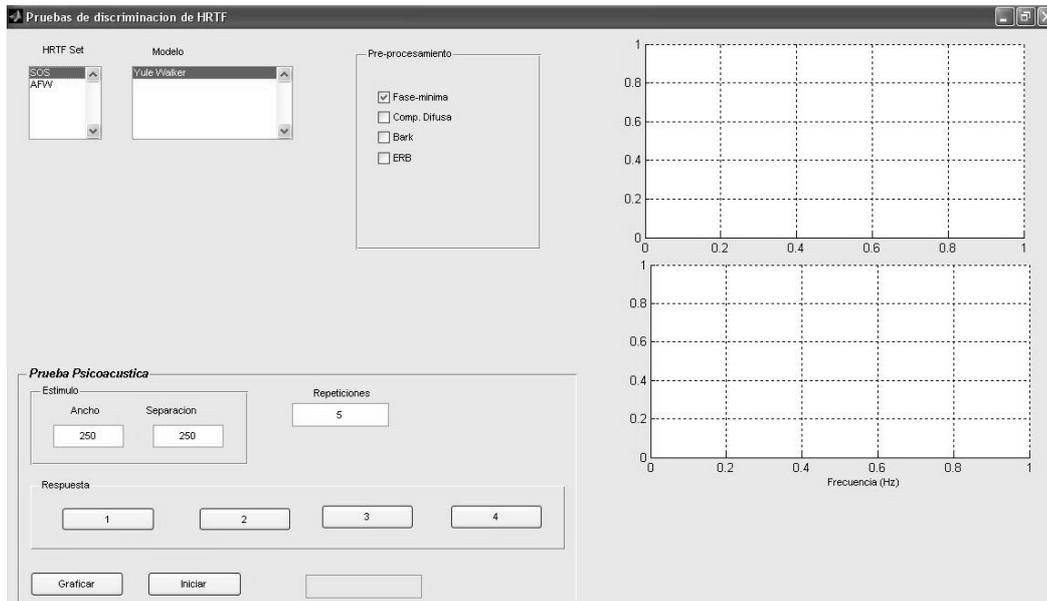


Figura 6.5 Software de administración de pruebas psicoacústicas que permite evaluar modelos del oyente de diferentes grados de complejidad

6.3.1 Estructura de alto nivel

En la Figura 6.6 se muestra un diagrama de la estructura de alto nivel del AVRS. El software del sistema se puede dividir en interfaz de configuración, procesos a nivel de usuario (tareas SRT) y procesos de RTAI-LXRT (tareas HRT).

La interfaz de configuración (Figura 6.7) fue creada para cargar los parámetros del sistema de una manera amigable. Posibilita determinar los valores de algunas variables (como por ejemplo, temperatura), configurar la geometría del recinto y los materiales de las superficies, realizar algunos cálculos (como el diseño de los filtros de absorción), determinar la posición y orientación de la fuente y el oyente, y realizar el cálculo de interpolación de las HRTFs.

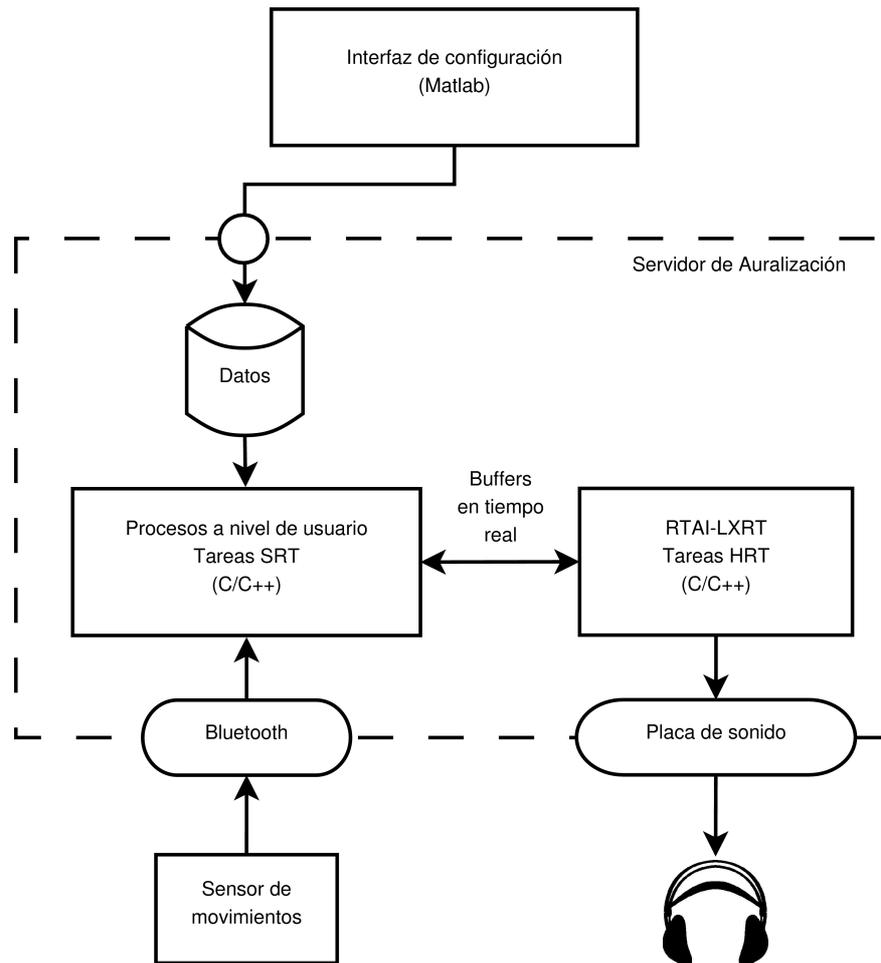


Figura 6.6 Estructura de alto nivel del AVRS

Está programada mediante scripts de MATLAB y se ejecuta en una computadora diferente al servidor de auralización (pero comunicada al mismo). Como no se ejecuta en tiempo real, la configuración resultante se guarda en el servidor de auralización¹⁹ para luego ser utilizada en la simulación.

¹⁹ No es posible modificar la configuración mientras se ejecuta la simulación.

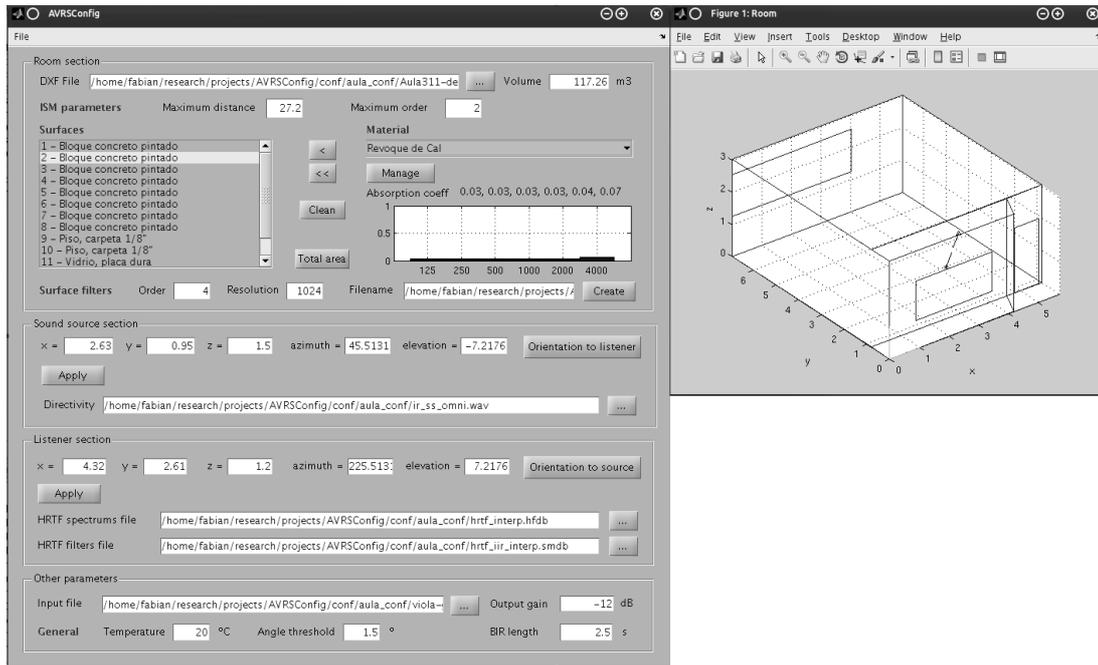


Figura 6.7 Interfaz de configuración del AVRS

Inicialmente, en el servidor, los procesos a nivel de usuario interpretan los parámetros de la línea de comandos y del archivo de configuración para llevar a cabo la simulación. Por ejemplo, para ejecutar el sistema con un determinado archivo de configuración, se hace²⁰

```
$ ./avrs <archivo de configuración>
```

Una vez en ejecución, las tareas SRT de los procesos a nivel de usuario controlan la entrada del usuario y capturan los datos que envía el sensor de movimientos de cabeza a través de la interfaz Bluetooth. Estos procesos se comunican con los procesos RTAI-LXRT mediante diferentes buffers en tiempo real. En las tareas HRT se realizan los cálculos de los modelos para lograr la

²⁰ Previamente se deben cargar los módulos RTAI correspondientes en el kernel.

síntesis de la respuesta impulsiva binaural (BIR), para luego realizar la convolución de baja latencia con una señal anecoica. El resultado de la simulación se puede escuchar mediante auriculares.

6.4 Ejecución en tiempo real

Durante su ejecución en tiempo real, el AVRS tiene un período aproximadamente 11,6 ms (512 muestras a 44,1 kHz). En ese tiempo el sistema actualiza la posición y orientación del oyente, realiza los cálculos de los modelos para obtener la BIR, convoluciona la BIR con una señal anecoica de entrada y reproduce la señal auralizada.

Todo esto se realiza siguiendo una estrategia de múltiples hilos y múltiples tareas, optimizando el uso del tiempo en los núcleos de la CPU. Cada subproceso se ejecuta en paralelo con el resto y el envío y recepción de datos se realiza de manera no bloqueante pero sincronizada. Es decir, no es necesario esperar a que un subproceso termine el envío de los resultados para que otro prosiga. Como se dijo, las tareas están comunicadas mediante diferentes tipos de buffers (RT-mailbox, RT-FIFOs, memoria compartida). A su vez, cuando hay una actualización del movimiento de la cabeza del oyente, se interrumpe el procesamiento actual para dar paso al nuevo cálculo inmediatamente.

La complejidad y precisión de los modelos implementados están solamente limitados a los recursos de hardware disponibles. Dado que el sistema es escalable, se puede actualizar el hardware de la computadora y, de esta manera, implementar modelos más complejos sin necesidad de cambiar la arquitectura del sistema. En la Figura 6.8 se puede observar el diagrama esquemático de los módulos que componen la arquitectura del AVRS.

Cuando se inicializa el sistema se cargan todos los datos necesarios para su ejecución en la memoria principal, éstos son: la geometría del recinto, la posición y orientación inicial de la fuente, las características de directividad de la fuente, la posición y orientación inicial del oyente, las características de los materiales, y las HRTFs. También se realiza la rutina de calibración del sensor de movimientos. Se le pide al sujeto que apunte su cabeza “al frente”. Esta orientación se corresponderá con la orientación inicial dada del oyente.

La síntesis binaural, es decir la construcción de la BIR del sistema, se divide en dos partes. La más importante es la primera, que contiene el sonido directo y las reflexiones tempranas. Cada una de estas reflexiones son representadas por fuentes virtuales, calculadas por el ISM. La segunda parte, reflexiones tardías, se sintetiza mediante FDN por única vez al comenzar la simulación, para minimizar el tiempo de cálculo, ya que la respuesta se mantiene constante mientras no cambie la geometría del recinto.

La señal anecoica a auralizar ingresa al sistema en forma de *stream*, la cual es convolucionada, mediante un motor de convolución de baja latencia (ver sección 6.4.4 más adelante), con la BIR. El sonido ya auralizado se reproduce mediante librerías de procesamiento de audio en tiempo real (RtAudio de STK) (ver sección 6.7.1 más adelante), utilizando un esquema de *callbacks*.

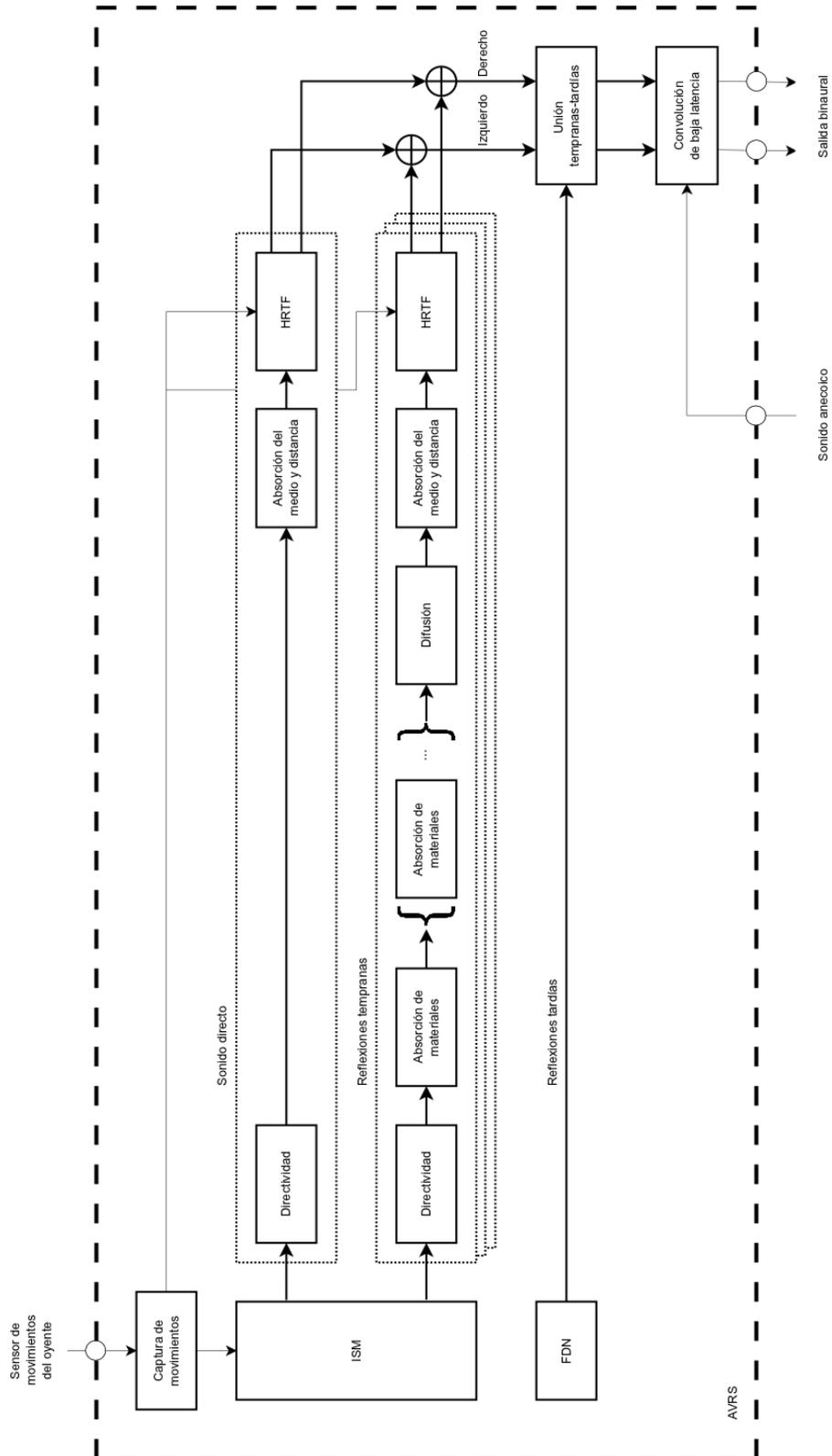


Figura 6.8 Diagrama esquemático de la arquitectura del AVRS

6.4.1 Actualización de los parámetros del sistema

Como se sabe, los parámetros del sistema cambian a medida que el sujeto interactúa con el entorno virtual. En un entorno dinámico e interactivo hay varias situaciones (eventos) que obligan a actualizar los parámetros para asegurar que la calidad no se degrade. En la Tabla 6.1 se muestran los eventos a tener en cuenta y se detallan las operaciones de cálculo que desencadenan cada uno de estos eventos en el ISM (reflexiones tempranas especulares).

Eventos	ISM		
	Recalcular posiciones	Validar audibilidad	Actualizar orientaciones
Movimiento de la fuente sonora	×	×	×
Giro de la fuente sonora			×
Movimiento de la cabeza del oyente		×	×
Giro de la cabeza del oyente			×

Tabla 6.1 Eventos a considerar en un sistema de simulación acústica virtual en tiempo real

Si la fuente sonora se mueve, todas las VSs se deben recalcular completamente. Cuando el oyente se mueve, se debe validar nuevamente la audibilidad de todas las fuentes virtuales. Como las posiciones de las fuentes virtuales no varían, no es necesario recalcularlas.

Si el oyente sólo gira la cabeza, o se gira la fuente, sin cambiar la posición, no hay cambios en la audibilidad de las fuentes virtuales y sólo se deben recalcular los ángulos de azimut y de elevación de las fuentes virtuales con respecto al oyente.

Actualmente, tanto el movimiento como el giro de la cabeza del oyente se tienen en cuenta en el AVRS.

6.4.2 Filtros para el sonido directo y las reflexiones tempranas

El AVRS requiere que todos los filtros se actualicen rápidamente, lo que es más desafiante aún al considerar el modelo del recinto. El procesamiento de la cadena de filtros para cada reflexión (incluido el sonido directo) de la primera parte de la RIR, es una parte crucial en el procesamiento en tiempo real.

Cada vez que el oyente realiza un movimiento, implica calcular nuevamente toda la cadena de filtros. Se computa la diferencia entre la última medición de la posición de la cabeza que superó el límite establecido (umbral) y la medición actual. Si dicho valor supera este umbral, se considera válido. A partir de evaluaciones realizadas con oyentes, los umbrales se establecieron en 2 cm para el movimiento de traslación y $1,5^\circ$ para un cambio de orientación de la cabeza²¹. Este procedimiento evita cálculos espurios o innecesarios como, por ejemplo, los que ocurrirían debido a pequeñas imprecisiones del sensor de movimientos de cabeza.

Como se pudo observar en la Figura 6.8, la cadena de filtros está compuesta por varias etapas. El primer filtro es el de directividad. Describe la dependencia de la dirección del sonido irradiado por la fuente sonora, es decir, define el peso que tiene cada reflexión en la síntesis binaural. Esto cobra más importancia en los sistemas dinámicos de auralización, donde no sólo el oyente se puede mover, sino también la fuente puede moverse y girar. Si bien, por el momento, el AVRS no permite que la fuente se mueva ni gire durante la simulación, está diseñado para que pueda incorporarse este aspecto en un futuro.

El AVRS acepta bases de datos de directividad de cualquier resolución espacial. Actualmente, las simulaciones se realizan con mediciones de una fuente sonora omnidireccional, descrita en el Capítulo 3. Se diseñaron filtros IIR de quinto

²¹ Estos valores son similares a los reportados por Lentz (2007).

orden para cada dirección que se almacena en memoria. También es posible construir los diagramas de directividad de una fuente en forma manual.

Los filtros de absorción de materiales, se diseñaron a partir de los respectivos coeficientes de absorción. Son filtros IIR de cuarto orden. 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 que generó la reflexión. Por ejemplo, si la reflexión se generó a partir de una VS de segundo orden, es decir, el rayo sonoro se reflejó en dos superficies, se aplicarán dos filtros de absorción de materiales (uno por cada material). Por el momento, son independientes del ángulo de incidencia del sonido. Esta simplificación es usual en los sistemas para modelado acústico de recintos, pero se podría extender estos filtros y considerar la dependencia del ángulo si fuese necesario.

Para simular las reflexiones difusas que se producen cuando una onda incide en una superficie rugosa, se aplicó el filtro de difusión planteado en el Capítulo 4. Es un filtro IIR de quinto orden y su respuesta se asemeja perceptualmente a la de la difusión. Se comprobó que el mismo efecto se puede lograr atenuando a la primera parte²² de la respuesta de la FDN con una función exponencial. Estas dos posibilidades están disponibles en el AVRS.

La absorción del sonido en el medio (aire) depende principalmente de la distancia, temperatura y humedad. Si bien se pueden ajustar filtros IIR de segundo orden a las respuestas en magnitud (Huopaniemi, Savioja, y Karjalainen 1997), actualmente se tiene en cuenta sólo la atenuación por la distancia (de acuerdo a la ley $1/r$) porque la influencia de los filtros es mínima y se economiza tiempo de cálculo.

²² Con un largo equivalente a las reflexiones tempranas.

Hasta aquí se manipuló una señal monoaural, pero el filtrado por HRTF genera una señal binaural. Son un par de filtros, uno para el oído izquierdo y el otro para el oído derecho, y el agregado del retardo correspondiente a la ITD. Así la señal binaural de salida adquiere toda la información direccional (ver Capítulo 5). Los filtros fueron diseñados por el método de Steiglitz-McBride con 5 iteraciones. Se trata de filtros IIR de orden 20.

Si bien existen otras formas más exactas para el cálculo de estos filtros, según estudios previos (Ramos, Araneda, y Tommasini 2009a; Tommasini y Ramos 2010), este método logra la mejor relación tiempo de cálculo–precisión si se compara con la convolución de las HRTFs en el dominio de las frecuencias o con simples filtros FIR.

El programa también acepta bases de datos de HRTFs de cualquier resolución. Para la tesis se utilizó la base de datos de HRIRs del CIPIC (Algazi et al. 2001) interpoladas a valores de azimut distanciados $1,25^\circ$ uno de otro, entre -80° a $+80^\circ$, y a valores de elevación distanciados $2,8125^\circ$ uno de otro, entre -45° y $+230,625^\circ$. Dando como resultado un total de 12771 posiciones (previamente se contaba con 1250) para cada oído, con 129 posiciones en azimut y 99 posiciones en elevación. Estos valores están en concordancia con el umbral MAA tanto para el plano horizontal (entre 1° y 2° al frente) (Mills 1958; Blauert 1997) como para el plano medio (3° a 9° en el hemisferio frontal) (Perrott y Saberi 1990; Blauert 1997).

Al finalizar la cadena de filtros, se deben combinar todos los resultados de las reflexiones tempranas en un buffer, para luego agregar la difusión (si se realiza mediante FDN) y las reflexiones tardías .

6.4.3 Generación de las reflexiones tardías

Las reflexiones tardías o cola de reverberación se asemejan a un ruido aleatorio

que se atenúa exponencialmente con el tiempo. Cuando la densidad de las reflexiones que alcanzan a un oyente es grande, éste no puede discriminar entre dos reflexiones sucesivas. Debido a estas evidencias, no es necesario modelar las reflexiones tardías de manera precisa como se hace con las tempranas especulares. Sin embargo, las reflexiones tardías están relacionadas con aspectos perceptuales particulares.

Como se vio en el Capítulo 4, utilizando la FDN de Jot con 8 líneas de retardo es posible simular las reflexiones tardías. Esta red permite alcanzar los tiempos de reverberación deseados. El problema es que el tiempo de preparación de la red (tiempo hasta que la salida se vuelve densa) es relativamente alto. Como este procesamiento se realiza “off-line”, se siguió la estrategia de realizar la preparación sin aplicar los filtros de atenuación. Luego de generar la salida para el doble de tiempo del retardo más largo ($\approx 51,2$ ms), ésta se vuelve lo suficientemente densa. Recién allí se comenzó a aplicar los filtros de atenuación y se descartó la parte anterior.

Para sintetizar en una sola BIR las reflexiones tempranas y las tardías, se utilizó el mixing time (Jot, Cerveau, y Warusfel 1997; Tommasini y Ramos 2008), que es una fórmula aproximada para establecer el tiempo de transición. Proviene de un análisis acústico-estadístico del campo sonoro en recintos. Con este valor, se calculó la distancia r que recorrió un rayo en dicho tiempo viajando a la velocidad del sonido. Esta distancia determinó la distancia máxima para el criterio de proximidad en el ISM. Es decir, sólo se generaron las VSs cuya distancia al oyente sea igual o menor a r . También se estimó la amplitud inicial de las reflexiones tardías con la ley $1/r$ (Tommasini y Ramos 2008) (Figura 6.9).

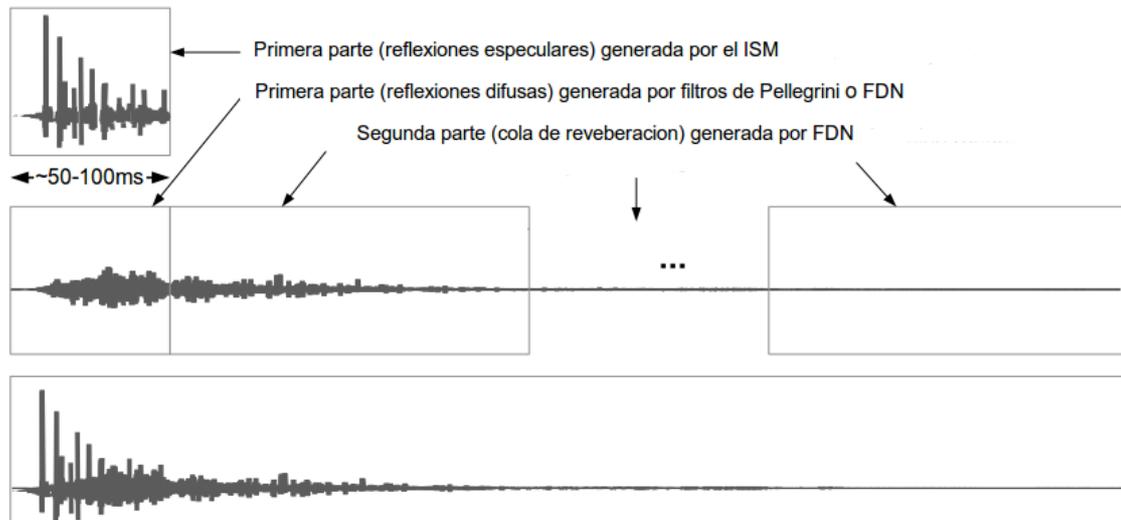


Figura 6.9 Combinación de las reflexiones tempranas con las reflexiones tardías

6.4.4 Convolución de baja latencia

La convolución de la señal anecoica de entrada con la BIR requiere de una gran potencia de cálculo. Cuando se realiza la auralización en campo libre, el tamaño del filtro (BIR) es de algunas cientos de muestras. Pero al incorporar el modelado acústico de recintos, esta cantidad se incrementa a quizás varios cientos de miles de muestras, dependiendo del tiempo de reverberación. Además, este procesamiento se debe realizar para cada oído²³.

Por ejemplo, una sala de concierto tiene un tiempo de reverberación que va desde los 1,8 s a 2,3 s, correspondientes a tamaños de filtros (BIR) de 80000 a 100000 muestras aproximadamente (para una frecuencia de muestreo de 44,1 kHz) (Lentz 2007). Un aula de clases, como la del caso de estudio que se presentará en el Capítulo 7, tiene un tiempo de reverberación de aproximadamente 2,5 s (\approx 110000 muestras), y el tiempo de reverberación de una iglesia o de salones muy

²³ Estos procesos se pueden realizar en paralelo, aunque esto no significa una disminución a la mitad del tiempo de cálculo.

grandes, a veces es superior a los 10 s (> 441000 muestras).

La carga computacional requerida para realizar todas las multiplicaciones y sumas involucradas en este tipo de cálculo es demasiado grande para una computadora estándar actual. Otra cuestión desfavorable es que la convolución causa latencias tan largas como el tamaño de la BIR. Por lo que el procesamiento en tiempo real se hace prácticamente imposible con estas técnicas, ya que el oyente no debe percibir estos retardos.

Para disminuir el tiempo de cálculo se puede realizar el proceso de convolución en el dominio de la frecuencia con algoritmos FFT (fast Fourier transform) optimizados. Si bien esto reduce el costo computacional al mínimo, la única manera de minimizar la latencia de la convolución es mediante un acondicionamiento especial en bloques de toda la BIR. De esta forma trabaja el método de convolución por particiones no uniformes (non-uniform partitioned convolution). El algoritmo procesa cada bloque individualmente en el dominio de la frecuencia. Una estrategia para el particionado eficiente es realizar una segmentación que vaya duplicando el tamaño del bloque cada dos bloques (Figura 6.10). Al principio se utilizan bloques pequeños para lograr una baja latencia de entrada-salida y, luego, se usan bloques cada vez más largos para ir reduciendo el costo computacional. Más detalles acerca de esta técnica de convolución se puede encontrar en Gardner (1995), Torger y Farina (2001), y Müller-Tomfelde (2001).

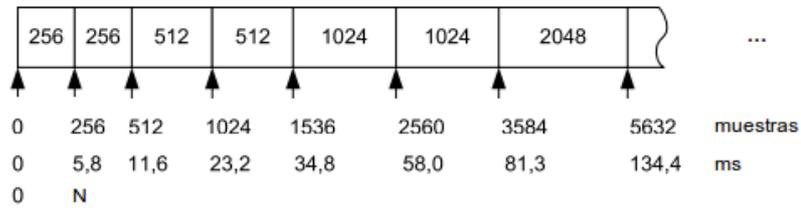


Figura 6.10 Acondicionamiento de los bloques para la convolución particionada no uniforme. Adaptado de Lentz (2007).

De esta manera se reduce el tiempo de cálculo y se logra un bajo retardo. La convolución por particiones no uniformes sólo agrega un retardo de $2N$, por lo que es recomendable usar bloques lo más pequeños posible ($N = 256$ muestras en esta tesis).

Además, el algoritmo implementado está preparado para hacer uso de estructuras de procesamiento específicas de hardware que incorporan los CPUs actuales, como SIMD (single instruction, multiple data)²⁴, disminuyendo aún más el tiempo de cálculo para algunas operaciones de cálculo vectorial.

6.5 Captura de los movimientos de cabeza

Con el objetivo de intensificar la experiencia del sujeto en un entorno virtual y asemejar dicha sensación a una situación real, es de suma importancia capturar los movimientos de cabeza del oyente. Por ejemplo, cuando una persona está inmersa en un campo sonoro, naturalmente tiende a girar su cabeza para identificar la dirección de un sonido (Arias et al. 2006; Ramos et al. 2009). Al mover la cabeza el escenario acústico virtual cambia y el sistema debe ser capaz de actualizarlo. Como se sabe, esto se debe realizar en un tiempo perceptualmente aceptable y con una

²⁴ Son instrucciones que aplican una misma operación sobre un conjunto relativamente grande de datos, para conseguir paralelismo en las operaciones.

rapidez que garantice cierta continuidad.

Existen diferentes alternativas comerciales para realizar la captura de los movimientos de la cabeza en tiempo real con seis grados de libertad (degrees of freedom, DOF)²⁵, es decir, los ángulos de rotación en los tres ejes (azimut, elevación y alabeo) y las coordenadas de la posición espacial. Estos dispositivos, como por ejemplo Polhemus G4²⁶, tienen un costo elevado y obligan a la utilización de software y librerías no libres. Por esto, se consideró la posibilidad de desarrollar un sensor de movimientos de bajo costo y basado en software libre.

En base a experiencias de trabajos anteriores (Lee 2008; Wronski 2008; Fleisch 2009; Riva 2009), se utilizó el controlador de la consola de juegos Nintendo Wii (conocido como wiimote) (Figura 6.11) para implementar un sensor de movimientos de cabeza. Su buena relación costo-prestaciones y la disponibilidad de librerías open-source fueron aspectos determinantes en la decisión. Es necesario aclarar que, como el sistema es flexible, escalable y de código abierto, con muy poco esfuerzo de programación es posible la incorporación de otro sensor.

Para sensar los movimientos de cabeza se utilizó la cámara digital del wiimote que permite identificar cuatro puntos de luz infrarroja dispuestos en un objeto de geometría conocida (beacon). El wiimote se comunica con la computadora enviándole las coordenadas de las proyecciones en el plano de la cámara de los cuatro puntos del beacon. Esta información es utilizada para realizar el cálculo de la posición relativa del beacon con respecto a la cámara.

A continuación se detallan los diferentes elementos que componen al sensor de movimientos de cabeza y su modo de funcionamiento.

25 Lo que permite la captura de los tres movimientos típicos de cabeza: rotación, tip y pivoteo (ver Capítulo 5).

26 www.polhemus.com/?page=Motion_G4



Figura 6.11 Diferentes vistas del wiimote. En la imagen de la derecha se puede ver la cámara del dispositivo.

6.5.1 Cámara del wiimote

La cámara del wiimote se compone de un sensor monocromático de 128×96 píxeles, un procesador de imágenes integrado y un filtro óptico infrarrojo, centrado en una longitud de onda de 940 nm. El procesador de imágenes integrado tiene la capacidad de efectuar el seguimiento de hasta cuatro puntos infrarrojos en movimiento. Usando un análisis subpíxel de $8\times$ provee una resolución de 1024×768 píxeles. Debido a que la velocidad de transmisión es limitada, sólo envía el tamaño de los puntos detectados y sus coordenadas en el plano del sensor monocromático, a una frecuencia de 100 Hz.

La cámara puede detectar cualquier fuente de luz infrarroja, esto incluye lámparas incandescentes, velas y luz solar. De acuerdo a Wronski (2008), las fuentes de luz infrarroja de una longitud de onda de 940 nm se detectan con el doble de intensidad que una fuente de 850 nm.

Al no existir información oficial del fabricante, algunos de los parámetros intrínsecos de la cámara, tales como la distancia focal, el campo de visión y el tamaño del píxel fueron obtenidos de mediciones realizados por Fleisch (2009) y

Wronski (2008). El resto de los parámetros fueron estimados mediante ensayos.

Se utilizó un valor de la distancia focal de 1380 píxeles, y el campo de visión resultó ser de 41° horizontalmente y 31° verticalmente.

6.5.2 Beacon

El beacon está compuesto por 4 LEDs infrarrojos de alta luminosidad de una longitud de onda de 940 nm. La apertura del haz es de 130° . El diseño del beacon se muestra en la Figura 6.12.

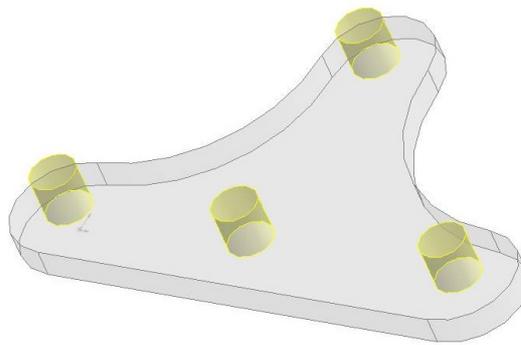


Figura 6.12 Diseño del beacon

Los LEDs están dispuestos de manera no coplanar, sobre una pieza de acrílico cuyas dimensiones se muestran en la Figura 6.13. Esta disposición es requerida por el algoritmo para calcular la posición relativa al origen de coordenadas (ubicado en el foco de la cámara) con 6DOF.

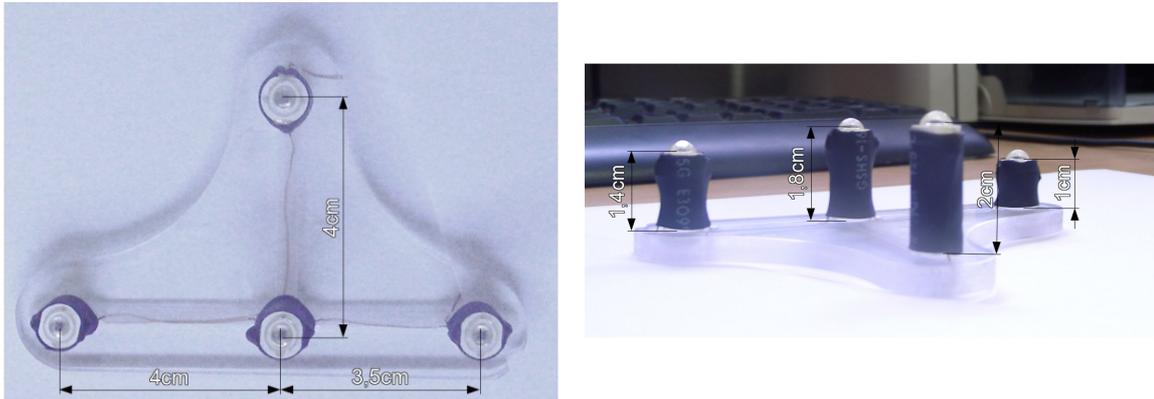


Figura 6.13 Medidas de la ubicación de los LEDs en el beacon

6.5.3 Montaje del sensor de movimientos de cabeza

Se montó el beacon sobre la vincha de sujeción de unos auriculares Sennheiser HD457 (Figura 6.14a). El wiimote se posicionó en un soporte montado sobre un trípode (Figura 6.14b) a aproximadamente 90 cm por encima del sujeto que porta los auriculares junto al beacon sobre su cabeza. En la Figura 6.15 se puede observar el montaje final con el AVRS funcionando.



(a)



(b)

Figura 6.14 Montaje del sensor de movimientos



Figura 6.15 Montaje final del AVRS

La distancia a la que se posicionó el wiimote y el tamaño del beacon se eligieron teniendo en cuenta los parámetros intrínsecos del wiimote y los parámetros de los LEDs, tales como potencia y directividad del haz. Si el beacon es muy pequeño, o si se posiciona muy lejos, el wiimote no puede discriminar entre dos puntos contiguos. Por el contrario, si el beacon es muy grande o está muy cerca de la cámara, se disminuye la zona de movilidad del sujeto (debido a la pérdida de visibilidad de algún punto infrarrojo).

6.5.4 Cálculo de la posición y orientación

Un módulo del AVRS, que utiliza la librería Cwiid (ver sección 6.7.1 más adelante) para la comunicación con el wiimote, usa los puntos captados por la cámara y la disposición de los LEDs infrarrojos para calcular la posición y orientación del beacon. Para su cálculo se realizan los siguientes pasos:

1. Se obtienen los cuatro puntos en 2D captados por la cámara, que son las proyecciones de los puntos del beacon en el plano de la cámara.
2. Se corresponden las proyecciones con sus respectivos puntos. El algoritmo relaciona las coordenadas en 2D de cada punto con las coordenadas 3D conocidas de los puntos infrarrojos del beacon.
3. Mediante el algoritmo POSIT (Pose from Orthography and Scaling with Iterations) (Dementhon y Davis 1995), se calcula la matriz de rotación y el vector de translación del beacon con respecto a la cámara del wiimote, El POSIT es un algoritmo iterativo que permite calcular la *pose*²⁷ 3D de un objeto de geometría conocida (beacon) a partir de los puntos 2D de una imagen (imagen de la cámara).
4. A partir de la matriz de rotación, se obtienen los ángulos de Euler. Los ángulos de Euler se pueden considerar como el resultado de tres rotaciones compuestas y las convenciones son nombradas de acuerdo a esta composición. Se ha utilizado la convención ZXZ.
De esta manera se representa la posición por las coordenadas (x, y, z) con respecto al origen ubicado en el foco de la cámara, y la orientación mediante los ángulos de Euler (θ, ϕ, ψ) (donde ψ no es de interés).
5. Finalmente, los resultados obtenidos son enviados a través de una estructura FIFO en tiempo real a las tareas HRT que utilizan esta información.

6.6 Sistemas de coordenadas en el AVRS

En el escenario virtual existen diferentes sistemas de coordenadas. El recinto

²⁷ En visión por computadora se le llama “pose” a la combinación de la posición y orientación de un objeto.

tiene el sistema de coordenadas \mathcal{R} cuyo inicio es ubicado arbitrariamente al modelarlo geoméricamente, el oyente posee el sistema de coordenadas \mathcal{L} cuyo inicio es el centro de la cabeza (ver Capítulo 5), y la fuente tiene el sistema de coordenadas \mathcal{S} cuyo inicio también se encuentra en el centro de dicha fuente.

Tanto la ubicación y posición de la fuente y el oyente están dados en \mathcal{R} . Pero las bases de datos de directividad y de HRTFs están dadas en \mathcal{S} y \mathcal{L} respectivamente.

Es de interés representar la ubicación de una fuente sonora (ya sea real o virtual), dada en el sistema \mathcal{R} , en el sistema de referencia \mathcal{L} . Esto permite hacer uso de la base de datos de HRTFs. Para lograrlo se utiliza el teorema de rotación de Euler, el cual indica que cualquier rotación se puede describir usando tres ángulos (θ, ϕ, ψ) (azimut, elevación y rol respectivamente). Si las rotaciones están dadas en términos de matrices \mathbf{A} , \mathbf{B} y \mathbf{C} , entonces la rotación general \mathbf{R} está dada por

$$\mathbf{R} = \mathbf{ABC} . \quad (6.1)$$

Como se dijo, se utiliza la convención ZXZ. En esta convención la primera rotación es de un ángulo θ sobre el eje z , la segunda es de un ángulo ϕ alrededor del eje x , y la tercera es por un ángulo ψ nuevamente sobre del eje z . Las matrices para esta convención son

$$\mathbf{A} = \begin{bmatrix} \cos \psi & \text{sen } \psi & 0 \\ -\text{sen } \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (6.2)$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \text{sen } \phi \\ 0 & -\text{sen } \phi & \cos \phi \end{bmatrix}, \quad (6.3)$$

$$\mathbf{C} = \begin{bmatrix} \cos \theta & \text{sen } \theta & 0 \\ -\text{sen } \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (6.4)$$

Como el ángulo de rol no se tiene en cuenta en la simulación, se hace $\psi = 0$ quedando $\mathbf{A} = \mathbf{I}$, donde \mathbf{I} es la matriz identidad. De acuerdo a la Ecuación (6.1), la matriz de rotación \mathbf{R} queda

$$\mathbf{R} = \mathbf{I}\mathbf{B}\mathbf{C} = \begin{bmatrix} \cos \theta & \text{sen } \theta & 0 \\ -\text{sen } \theta \cos \phi & \cos \theta \cos \phi & \text{sen } \phi \\ \text{sen } \theta \text{sen } \phi & -\cos \theta \text{sen } \phi & \cos \phi \end{bmatrix}. \quad (6.5)$$

Siendo $[x, y, z]_{\mathcal{R}}^T$ la ubicación de la fuente sonora en el sistema \mathcal{R} , para representarla en el sistema de coordenadas \mathcal{L} del oyente, se hace

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathcal{L}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathcal{R}} \mathbf{R}. \quad (6.6)$$

De una manera similar se procede para representar la ubicación de las reflexiones (sistema \mathcal{R}) en el sistema de coordenadas \mathcal{L} . Así se hace uso de la base de datos de patrones de directividad.

6.7 Software y hardware utilizados

6.7.1 Dependencias de software

Para ejecutar el AVRS se utiliza un RTOS con el kernel GNU/Linux versión 2.6.28.9 modificado con la extensión para tiempo real RTAI 3.7.1. Además, se requieren las siguientes librerías open source para su funcionamiento:

- Armadillo 1.2.0. Librería C++ para cálculos de álgebra lineal (Sanderson 2010).
- FFTW 3.2.2. Librería C para calcular la transformada discreta de Fourier (discrete Fourier transform, DFT) mediante algoritmos FFT optimizados

(Frigo y Johnson 2005).

- STK 4.4.2. Conjunto de clases C++ para síntesis y procesamiento de señales (Scavone y Cook 2005).
- dxflib 2.2.0.0²⁸. Librería C++ para interpretar los archivos DXF con la geometría de los recintos.
- CWiid 0.6.00²⁹. Librería C para controlar el wiimote, usado por el sensor de movimientos de cabeza.

También fueron incorporadas al sistema las siguientes cabeceras (headers) y clases open-source:

- Clase C++ de convolución con baja latencia, la cual fue adaptada del sistema SSR (Ahrens, Geier, y Spors 2008).
- Clase C++ tree.hh³⁰ versión 2.65 desarrollada por Kasper Peeters. Utilizada para la construcción de la estructura tipo árbol para el ISM.
- Clase C++ ConfigFile³¹ versión 2.1 desarrollada por Rick Wagner, adaptada para leer los archivos de configuración.

6.7.2 Hardware

En términos de hardware, el AVRS está compuesto por el servidor de auralización, un sensor de movimientos de cabeza y unos auriculares. El servidor, donde se ejecuta el software en tiempo real, consta de un procesador Intel Core 2 Quad Q8200 (arquitectura x86-64 de 4 núcleos), 4 GB de memoria RAM DDR2 800 MHz y una placa de sonido Intel HDA con codec Realtek ALC888.

28 www.ribbonsoft.com/dxflib.html

29 abstrakraft.org/cwiid

30 tree.phi-sci.com

31 www-personal.umich.edu/~wagnerr/ConfigFile.html

Para obtener la posición y orientación de la cabeza del sujeto se utiliza el sensor de movimientos de bajo costo mostrado anteriormente en la sección 6.5. Para la reproducción de los sonidos sintetizados se utiliza auriculares Sennheiser HD457.

Debido a que el hardware es estándar, cualquier componente puede ser reemplazado y actualizado sin afectar al resto. Esto le permite al sistema ser completamente escalable.

7 – Resultados de un caso de estudio

Para evaluar y validar el AVRS, se realizó la simulación de un aula de clases. Tomando a este recinto como caso de estudio para todo el capítulo, a continuación se muestran los resultados vinculados a la evaluación del comportamiento y el rendimiento de los diferentes modelos implementados en el sistema. Un estudio comparativo con mediciones de RIRs efectuadas en el aula permitió verificar el grado de ajuste de la simulación. Además, se realizó la validación del ajuste de las HRTFs modeladas. Como se mencionó en capítulos anteriores, el objetivo del AVRS es lograr que los errores se encuentren por debajo de las jnd correspondientes, esto es, la mínima diferencia que una persona es capaz de discriminar de un parámetro físico-acústico en particular. Como en todo sistema de realidad acústica virtual, además, fue importante realizar un estudio detallado de los tiempos de cálculo y de las latencias.

7.1 Presentación del caso

El escenario elegido para el caso de estudio es un aula de clases de la Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN, FRC). El

recinto se encuentra ubicado en el Edificio Malvinas de dicha facultad³².

En la Figura 7.1 se muestra un *render* del recinto, que es un monoambiente de 117,26 m³ de volumen. Sus dimensiones interiores máximas son: 5,74 m de ancho 6,93 m de largo y 3,03 m de alto. La superficie total interior es de 156,32 m² (incluyendo piso, paredes y techo).

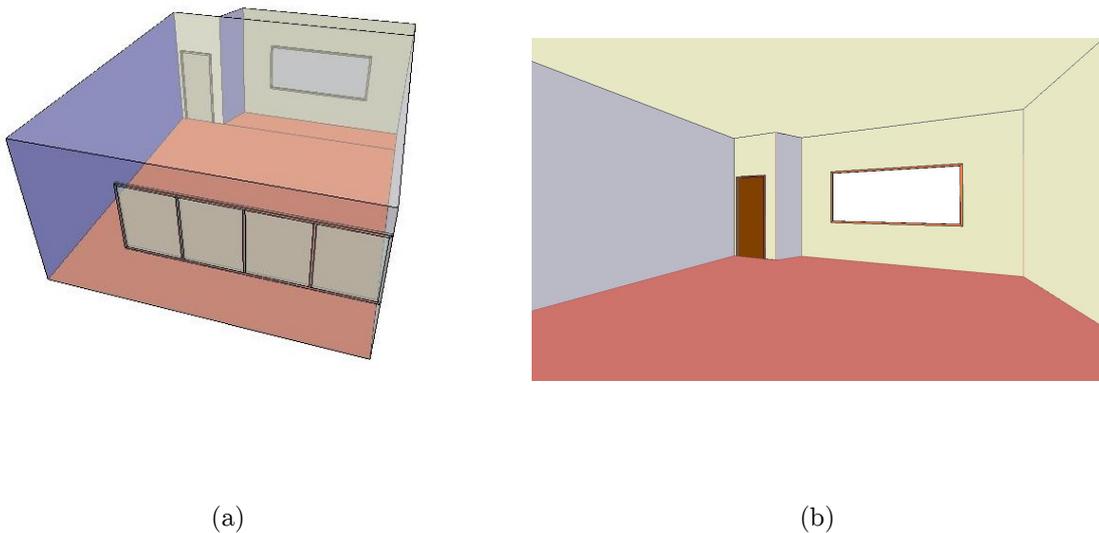


Figura 7.1 Render del recinto. a) Vista exterior, y b) vista interior.

Las paredes están construidas con ladrillo macizo de barro cocido de 0,15 m de espesor, revocadas y pintadas. Sobre una de las paredes se encuentra un pizarrón de vidrio con marco metálico y una puerta de ingreso de madera. En la pared opuesta hay una ventana con vidrios de 4 mm de espesor que ocupa prácticamente todo el ancho de la pared.

El techo es de losa de hormigón de aproximadamente 0,20 m de espesor, revocado y pintado. El piso es de baldosas de granito.

³² Actualmente dicha aula ha sufrido modificaciones debido a refacciones realizadas en el edificio.

7.2 Medición acústica del recinto

Se midió la RIR según los lineamientos planteados en la norma ISO 3382 (2009) con la finalidad de comparar la verosimilitud de la RIR generada por el modelo del recinto (Cravero et al. 2011).

7.2.1 Fuente sonora

La fuente utilizada para la medición fue una fuente sonora omnidireccional (OSS) tipo dodecaedro que cumple con la norma especificada (ver Capítulo 3 para más detalles de la OSS). Su ancho de banda se extiende de 31,5 Hz a 8 kHz.

7.2.2 Señal de excitación

Históricamente, para obtener la RIR, se utilizaba como señal de excitación ráfagas de ruido filtradas por bandas o disparos de pistola (Schroeder 1965). La desventaja de estos métodos es la pobre SNR debido a que su energía está concentrada en un pequeño intervalo de tiempo.

En la actualidad, es habitual utilizar señales continuas que mediante un procesamiento a posteriori permiten obtener la RIR con una mejor SNR (Vanderkooy 1994; Daigle y Xiang 2006). Estas técnicas, conocidas como métodos indirectos, han sido adoptadas por la normativa ISO 3382 (2009), específica del estudio de recintos. Por otra parte, se ha detectado que existen variaciones en los resultados según la técnica empleada (Vorländer y Bietz 1994; Ferreyra y Ramos 2007).

Para este estudio se utilizó el método de respuesta impulsiva integrada basado en la deconvolución de señales determinísticas (barridos en frecuencia de una señal sinusoidal que se incrementa exponencialmente con el tiempo, e-sweep) (Farina

2000), con una señal de 23,8 s de duración³³. Actualmente, este método presenta las mejores características de SNR, discriminación y eliminación de distorsiones no lineales introducidas por el sistema de medición, repetibilidad y precisión.

7.2.3 Equipamiento utilizado

En la Tabla 7.1 se detalla el listado completo del equipamiento utilizado para la medición.

Descripción	Marca/modelo
Placa de adquisición de sonido 44,1 kHz, 16 bits	Realtek, HD Audio
Micrófono electrostático omnidireccional (MIC)	Brüel & Kjær, 4133
Preamplificador de micrófono, 200 V	Brüel & Kjær, 2804
Software de aplicación	Brüel & Kjær type 7481 (Dirac v3.0)
Consola de mezcla 12 Canales	Behringer, Eurorack UB1204FXpro
Fuente sonora omnidireccional (OSS)	CINTRA, OSS-12
Amplificador de potencia	Brüel & Kjær type 4224
Distanciómetro láser	HD spectra 150
Medidor de nivel sonoro	Brüel & Kjær 2250

Tabla 7.1 Equipamiento utilizado en la medición de la RIR

7.2.4 Punto de medición

Siguiendo las recomendaciones que plantea la norma ISO 3382, se seleccionaron nueve posiciones de medición (Figura 7.2). Este caso de estudio se centrará sobre una de estas posiciones, la posición M5 mostrada en la figura.

³³ El método está descrito en la norma ISO 3382.

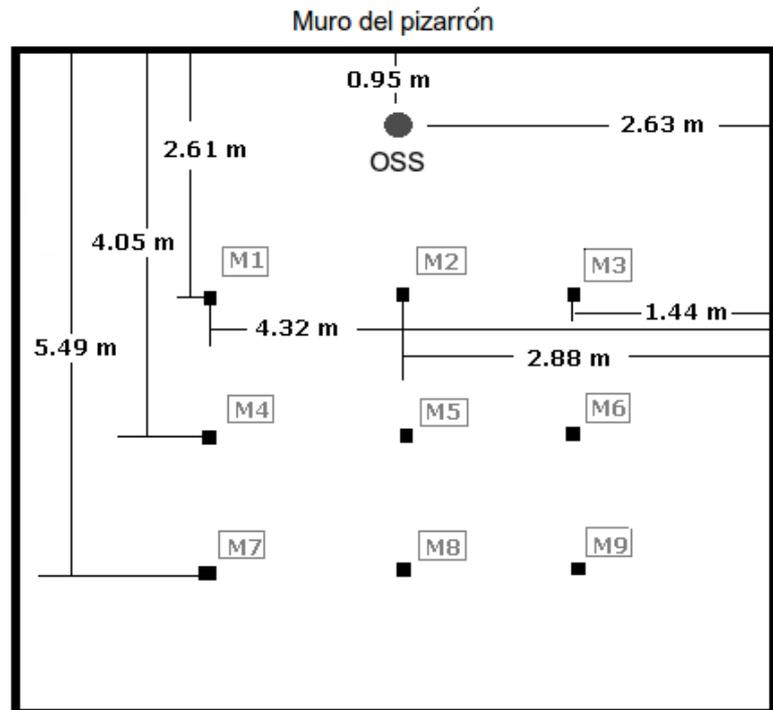


Figura 7.2 Puntos de medición. Para el caso de estudio se eligió el punto M5.

La medición se realizó con el aula vacía, libre de bancos y de personas. La altura de la OSS (centro del dodecaedro) es de 1,5 m de altura y la del micrófono es de 1,2 m (altura aproximada de una persona sentada). En la Figura 7.3 se muestran fotografías tomadas al momento de la medición.



Figura 7.3: Fotografías del aula al momento de la medición

7.2.5 Resultados de la medición

La RIR medida en el punto M5 se muestra en la Figura 7.4.

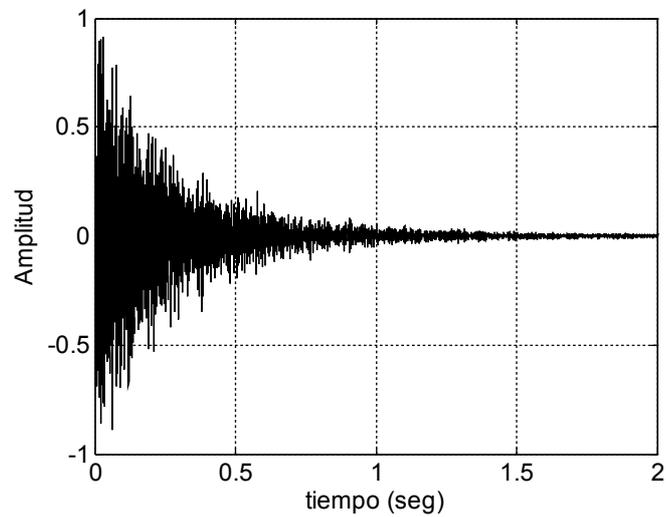


Figura 7.4 RIR medida según norma ISO 3382

Como se vio en el Capítulo 4, a partir de la RIR se pueden calcular un conjunto de parámetros que ponderan diversas características acústicas del recinto

(ver Apéndice A). Estos parámetros están relacionados con juicios subjetivos concernientes a la calidad acústica. Para este tipo de recinto, las bandas de octava relevantes van desde los 250 Hz³⁴ hasta los 4 kHz (Tang y Yeung 2006). En la Tabla 7.2 se muestran los valores de dichos parámetros para estas frecuencias de la RIR medida³⁵.

Parámetro	Frecuencia [Hz]				
	250	500	1000	2000	4000
EDT [s]	2,96	2,34	1,95	1,66	1,44
T ₁₀ [s]	3,14	2,48	1,94	1,66	1,50
T ₂₀ [s]	3,34	2,38	1,89	1,77	1,50
T ₃₀ [s]	3,39	2,46	2,08	1,82	1,52
C ₅₀ [dB]	-4,28	-6,16	-3,80	-2,81	-2,14
C ₈₀ [dB]	-3,21	-1,32	-1,68	0,15	0,90
D ₅₀ [-]	0,27	0,20	0,29	0,34	0,38
Ts [ms]	203,11	158,88	145,19	119,95	102,52

Tabla 7.2 Parámetros característicos de la RIR medida

7.3 Simulación acústica del recinto

Simulando el recinto descrito, se obtuvieron los resultados que se detallan a continuación.

7.3.1 Modelo de la fuente sonora

Para que los resultados sean comparables, se utilizaron las mediciones de

³⁴ La frecuencia de Schroeder (que indica la frecuencia a partir de la cual la simulación es válida) es cercana a los 300 Hz para este caso (ver Capítulo 4).

³⁵ Todos estos parámetros fueron calculados con un programa desarrollado especialmente que permite comparar RIRs medidas y simuladas, el cual ya se mencionó en el Capítulo 6.

directividad de la OSS del CINTRA. A partir de las cuales se diseñaron los filtros.

7.3.2 Modelo del recinto

En primer lugar se construyó un modelo CAD del aula, para luego convertirlo en un modelo geométrico con información de la fuente y del oyente, como así también de los materiales de construcción. Además de las paredes, techo y piso, se incluyó el pizarrón, la puerta de ingreso y la ventana (Figura 7.5).

En el modelo geométrico se realizaron las simplificaciones sugeridas por Vorländer (2007). Los objetos y las ondulaciones superficiales que son pequeñas en comparación con las longitudes de onda (resolución de 0,5 m), fueron reemplazadas por superficies planas con las propiedades acústicas adecuadas.

A cada superficie se le asignó el material correspondiente, con sus coeficientes de absorción. Es necesario considerar cuidadosamente los coeficientes de absorción a utilizar. Se debe disponer de coeficientes fiables, ya que se comprobó que una mínima diferencia en sus valores puede acarrear errores en la simulación.

Luego, se simuló la RIR utilizando VSs de hasta segundo orden en el ISM, resultando en un total de 98. De éstas, sólo 28 son audibles para dichas posiciones de fuente y de oyente (Figura 7.6).

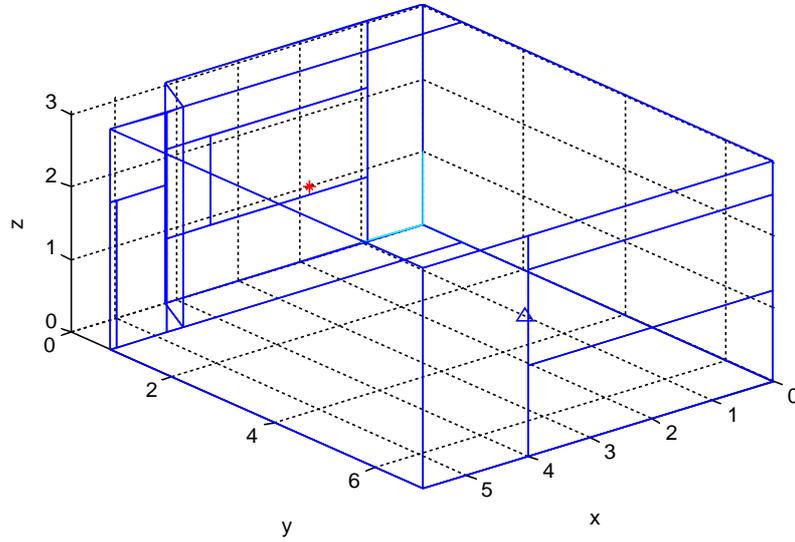


Figura 7.5 Posición de la fuente sonora (asterisco) y del oyente (triángulo)

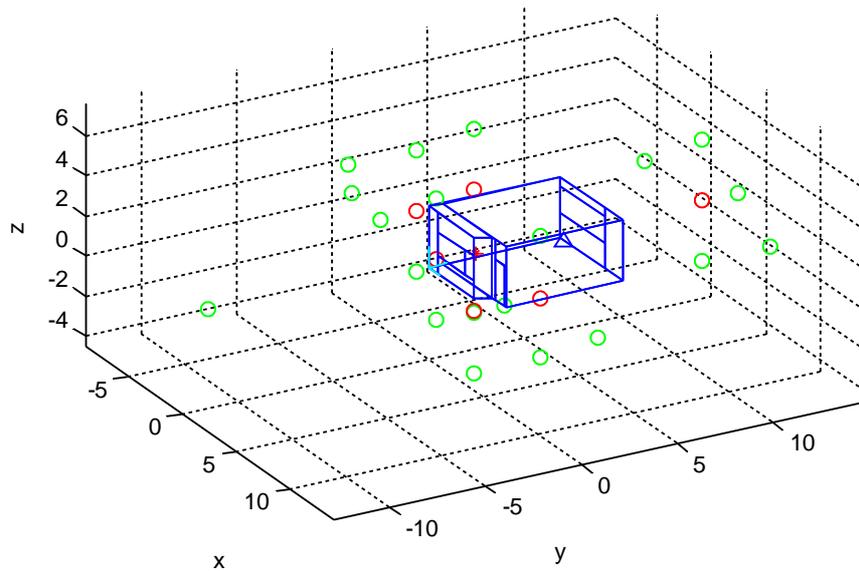


Figura 7.6 Distribución espacial de las VSs de primer y segundo orden

Para ejemplificar las reflexiones tempranas, en la Figura 7.7, se muestra la trayectoria de los rayos sonoros para una VS de primer orden originada en el techo y una de segundo orden, producida por reflexiones en la pared detrás de la fuente y en una de las paredes laterales.

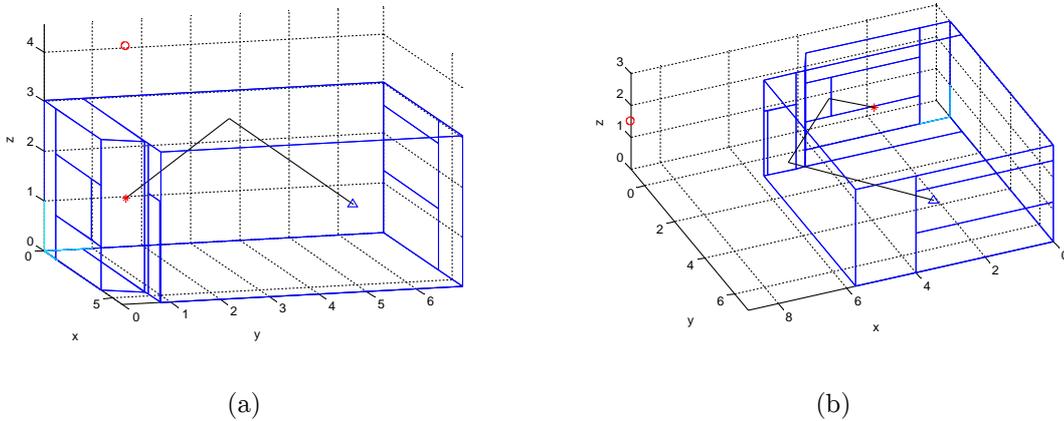


Figura 7.7 Trayectoria de los rayos para (a) VS de primer orden, y (b) VS de segundo orden

7.3.3 Resultados de la simulación

En la Figura 7.8 se muestra la RIR simulada en el dominio del tiempo. Se puede comparar con la Figura 7.4 para verificar que, si bien son similares, no son exactamente iguales. Lo importante es que las diferencias se encuentren por debajo de los límites perceptuales (ver siguiente sección).

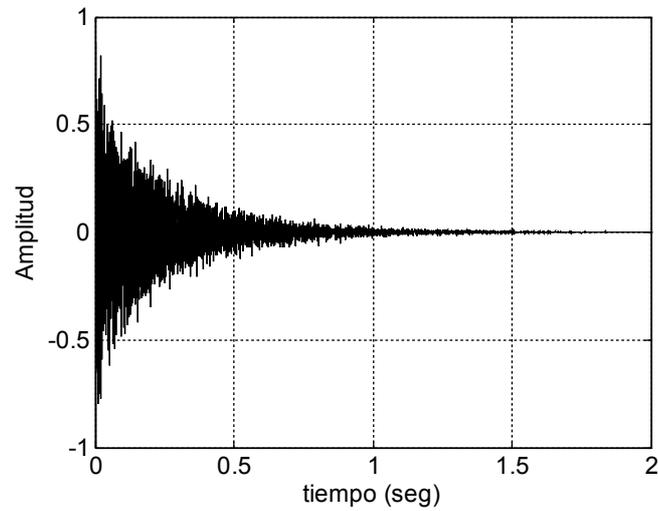


Figura 7.8 Respuesta impulsiva obtenida de la simulación

En la Tabla 7.3 se muestran los valores de los parámetros calculados a partir de la simulación para las bandas de frecuencia mencionadas.

Parámetro	Frecuencia [Hz]				
	250	500	1000	2000	4000
EDT [s]	2,91	2,51	2,09	1,83	1,44
T ₁₀ [s]	2,91	2,53	2,10	1,84	1,47
T ₂₀ [s]	2,91	2,54	2,11	1,86	1,48
T ₃₀ [s]	2,91	2,56	2,11	1,86	1,50
C ₅₀ [dB]	-5,03	-5,18	-3,35	-2,88	-0,52
C ₈₀ [dB]	-2,69	-2,28	-0,55	-0,23	1,86
D ₅₀ [-]	0,24	0,23	0,32	0,34	0,47
Ts [ms]	204,29	176,58	141,18	126,50	91,24

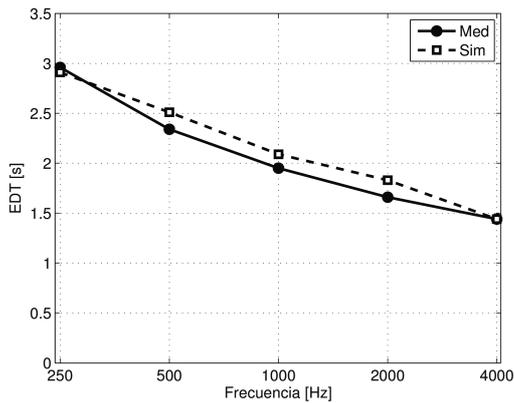
Tabla 7.3 Parámetros característicos de la RIR simulada

7.4 Validación de la simulación acústica del recinto

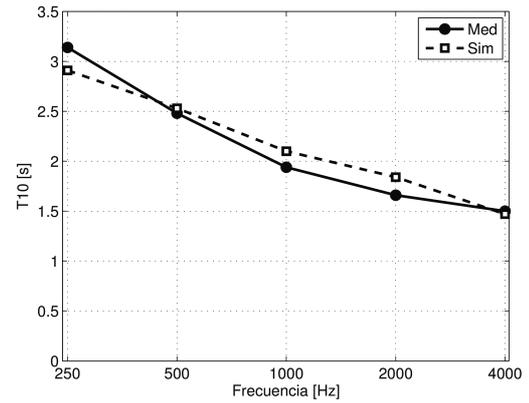
La simulación realizada en la sección anterior incluye el modelo de la fuente y

el modelo del recinto. A continuación se realiza una validación objetiva.

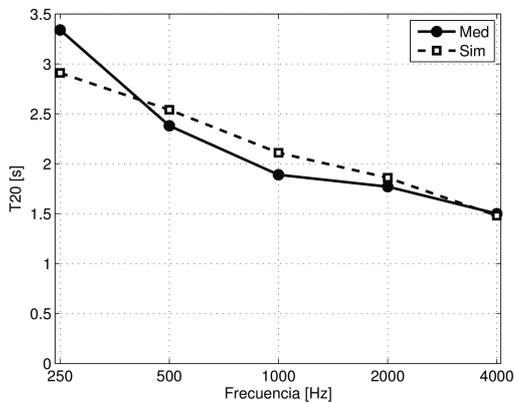
En la Figura 7.9 se muestran los tiempos de reverberación (EDT , T_{10} , T_{20} y T_{30}) calculados a partir de la RIR medida y simulada por bandas de una octava (entre 250 Hz y 4 kHz).



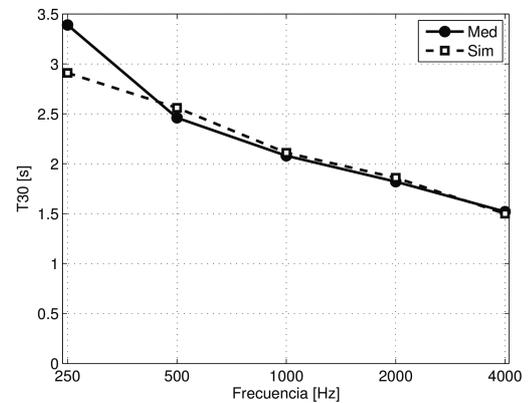
(a)



(b)



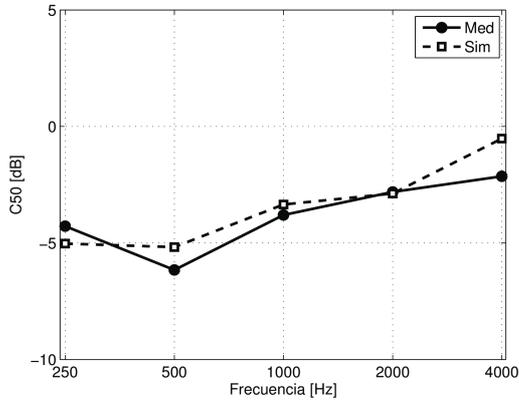
(c)



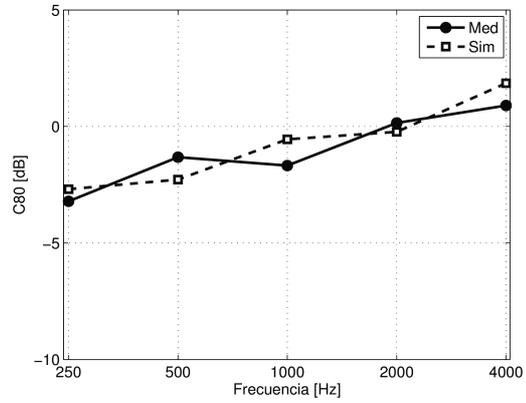
(d)

Figura 7.9 Gráficos de (a) EDT , (b) T_{10} , (c) T_{20} y (d) T_{30} obtenidos de la RIR medida (línea continua) y de la RIR simulada (línea a trazos)

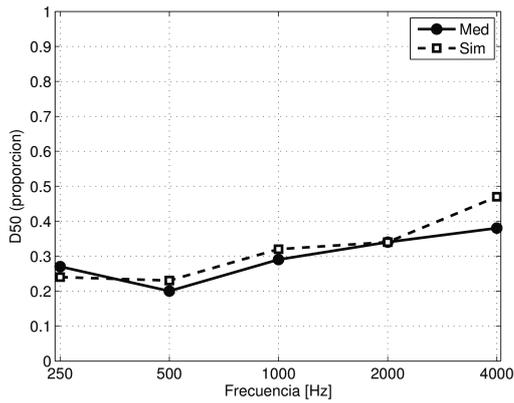
Los parámetros relacionados con la distribución de energía entre reflexiones tempranas y tardías, se grafican en la Figura 7.10.



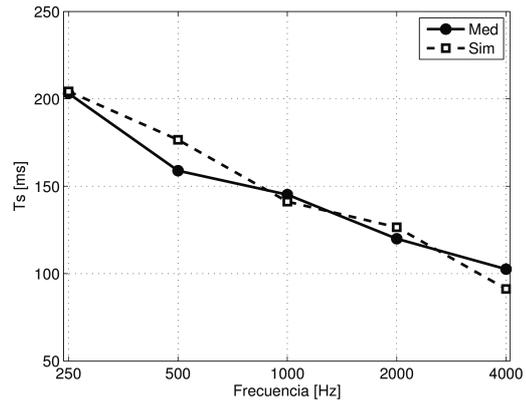
(a)



(b)



(c)



(d)

Figura 7.10 Gráficos de (a) C_{50} , (b) C_{80} , (c) D_{50} y (d) T_s obtenidos de la RIR medida (línea continua) y de la RIR simulada (línea a trazos)

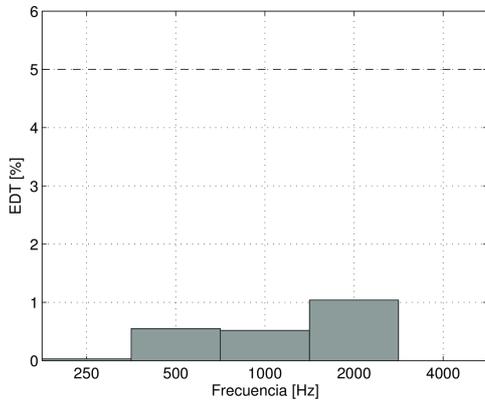
En la Tabla 7.4 se muestran los porcentajes del error medio cuadrático para los diferentes tiempos de reverberación y para el parámetro D_{50} , y los valores absolutos del error para los parámetros restantes. En la columna de la derecha se encuentran las j_{nd} para cada uno de los parámetros recomendados por Martellotta (2010).

Parámetro	Frecuencia [Hz]					jnd
	250	500	1000	2000	4000	
EDT [%]	0,03	0,55	0,52	1,04	0,00	5%
T ₁₀ [%]	0,52	0,03	0,65	1,17	0,05	5%
T ₂₀ [%]	1,64	0,47	1,24	0,25	0,02	5%
T ₃₀ [%]	2,02	0,17	0,02	0,05	0,02	5%
C ₅₀ [dB]	0,75	0,98	0,44	0,07	1,62	1 dB
C ₈₀ [dB]	0,52	0,96	1,13	0,38	0,96	1 dB
D ₅₀ [%]	1,47	3,80	0,56	0,01	5,77	5%
Ts [ms]	1,18	17,70	4,01	6,56	11,28	10 ms

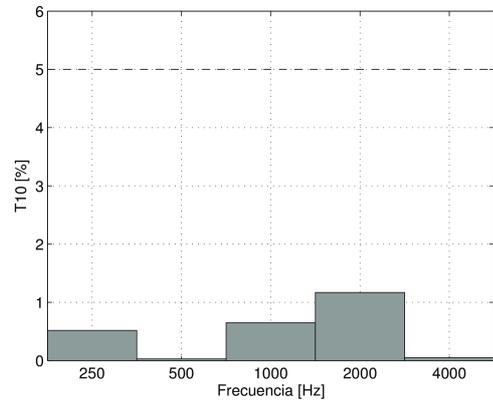
Tabla 7.4 Errores calculados entre la RIR medida y la simulada, junto a la jnd correspondiente

Como se puede observar, los valores obtenidos están por debajo de las jnd, (salvo aquellos resaltados en negrita). No se han encontrado en la literatura otros trabajos que validen los resultados de la simulación de recintos en tiempo real utilizando evidencias perceptuales.

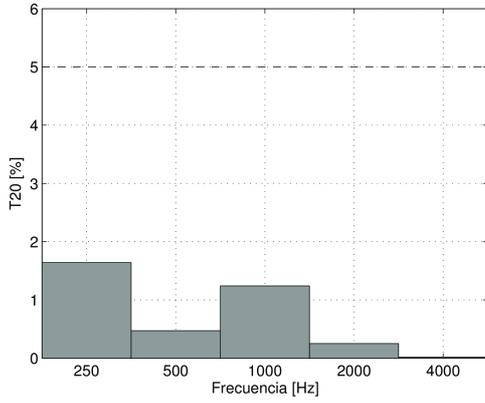
En la Figura 7.11 se muestra los histogramas de errores por frecuencia y la jnd recomendada (línea a trazos) para los parámetros de tiempo de reverberación. De la misma manera, en la Figura 7.12 se muestra los histogramas de errores para los parámetros relacionados con la distribución de energía entre reflexiones tempranas y tardías.



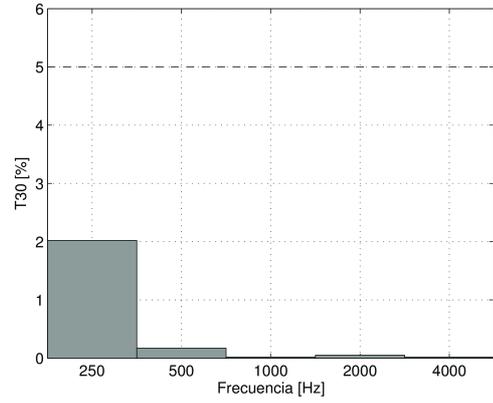
(a)



(b)



(c)



(d)

Figura 7.11 Histograma de errores por frecuencia para (a) EDT, (b) T_{10} , (c) T_{20} y (d) T_{30} . Las jnd se muestran en línea a trazos.

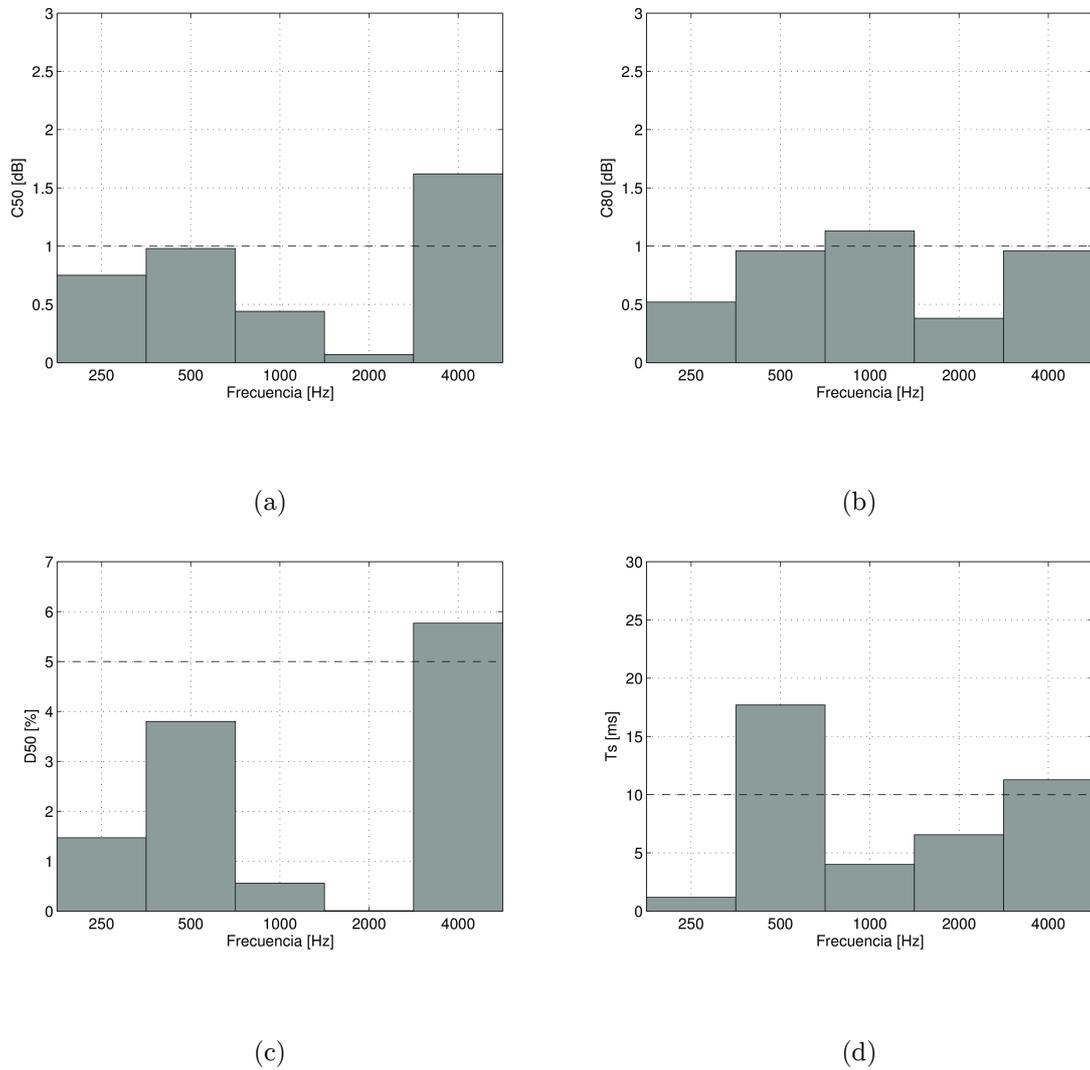


Figura 7.12 Histograma de errores por frecuencia para (a) C_{50} , (b) C_{80} , (c) D_{50} y (d) T_s . Las jnd se muestran en línea a trazos.

Actualmente, diferentes investigaciones evidencian la dependencia de las jnd de algunos parámetros con el tiempo de reverberación, como es el caso de T_s . Pero otros, como por ejemplo el C_{80} , resultan independientes (Martellotta 2010).

El estudio realizado por Ahearn et al. (2009), sobre la jnd del parámetro C_{80} , concluye que, para un total de 51 sujetos con entrenamiento auditivo, la jnd puede

alcanzar valores de hasta $1,6 \text{ dB} \pm 0,1 \text{ dB}$ en condiciones específicas. En el trabajo realizado por Okano (2002) se analizan y relacionan parámetros objetivos y subjetivos de recintos estableciendo valores para algunas jnd en salas de concierto.

Hasta donde se ha podido indagar, el estudio y determinación de las jnd para salas de estudio y aprendizaje (como es el caso de las aulas), donde la voz hablada es la principal señal de excitación, aún no han sido definidas y validadas.

7.5 Validación del modelo del oyente

No es trivial el problema de modelar al oyente mediante filtros diseñados a partir de las HRIRs. Si bien es posible utilizar técnicas tradicionales de diseños de filtros, la validación de la calidad del ajuste requiere de consideraciones particulares. La diferencia de tiempo interaural (ITD) tiene un rol dominante en bajas frecuencias (debajo de 1,5 kHz), mientras que la diferencia de nivel interaural (ILD) hace lo propio para altas frecuencias (Blauert 1997). Se sabe que en lugares del espacio donde la ITD y la ILD son mínimas o nulas (plano medio) o ambiguas (cono de confusión), los perfiles espectrales cumplen un papel fundamental en todo el rango de frecuencias (Wightman y Kistler 1989a; Wightman y Kistler 1989b). Ahora, la pregunta es cuán exactas deben ser modeladas estas claves para producir una imagen auditiva convincente. Afortunadamente, numerosos estudios psicoacústicos sobre la audición espacial han intentado (y continúan intentando) responder a esta pregunta. Diversos son los factores que contribuyen a una verídica percepción de una fuente sonora simulada con auriculares.

En esta sección, se valida objetivamente el modelo propuesto para el oyente haciendo hincapié en evidencias perceptuales de la audición espacial.

7.5.1 Base de datos de HRTFs

Como ya se había adelantado, se utilizó la base de datos del CIPIC (Center for Image Processing and Integrated Computing) de la Universidad de California. Consta de HRIRs (dominio del tiempo) medidas a las entradas de los conductos auditivos bloqueados de 45 sujetos para 1250 posiciones de la fuente sonora (Figura 7.13). Las HRIRs son secuencias de 200 puntos muestreadas a 44100 Hz y están compensadas en campo libre. La ubicación de la fuente se especifica por el ángulo de azimut θ (25 ángulos diferentes) y el ángulo de elevación ϕ (50 ángulos diferentes) referidos a un eje que pasa por ambos oídos: el eje interaural (ver Apéndice B) (Algazi et al. 2001).

Las mediciones se realizaron para elevaciones que varían uniformemente en pasos de $5,625^\circ$ entre -45° y $+230,625^\circ$. Los ángulos de azimut son: -80° , -65° , -55° ; $+55^\circ$, $+65^\circ$ y $+80^\circ$ y en pasos de 5° entre -45° y $+45^\circ$. Azimut 0° y elevación 0° corresponde al frente del sujeto; azimut 0° y elevación 180° detrás del sujeto; ángulos de azimut negativos corresponde a la izquierda y ángulos positivos a la derecha del sujeto; elevaciones negativas y positivas mayores a 180° debajo del eje interaural, enfrente y detrás respectivamente.

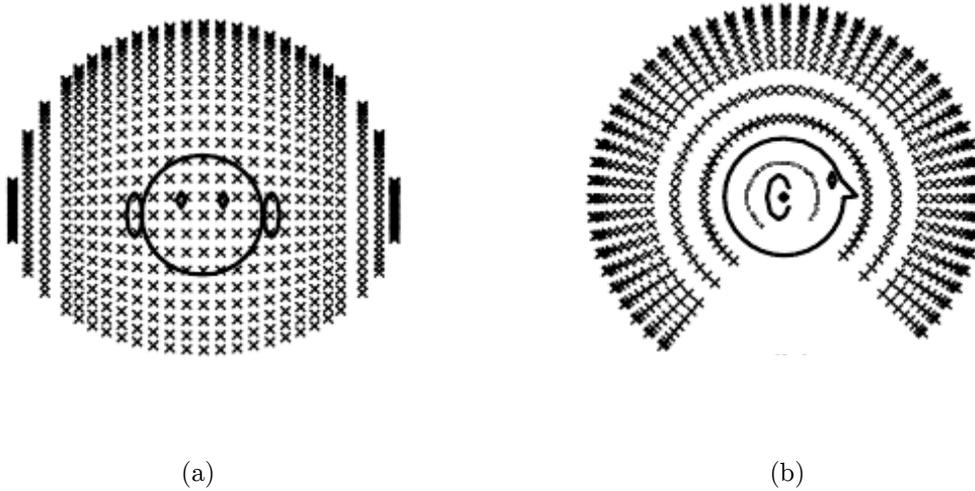


Figura 7.13 Ubicación de los puntos de medición: (a) vista frontal, y (b) vista lateral. Extraído de Algazi et al. (2001).

Además de las HRIRs, la base de datos cuenta con mediciones de algunos de los rasgos anatómicos de los sujetos, que se relacionan con la conformación de las HRIRs. Se trabajó con un subgrupo de 35 sujetos, que son los que cuentan con la totalidad de las mediciones antropométricas mencionadas.

7.5.2 Procedimiento

Previo al diseño de los filtros IIR por el método de Steglitz-McBride (ver Capítulo 5), se le quitaron los retardos iniciales al conjunto de las HRIRs. Estos retardos se deben, por un lado, a la distancia entre la fuente y los micrófonos (común a ambos oídos), y por el otro, a las ITDs³⁶.

Luego de realizar la interpolación espacial, se calcularon los coeficientes de los filtros de orden 20 (20 polos y 20 ceros) por el método propuesto. Se trabajó en el rango de 0 a 15000 Hz, ya es allí donde están concentradas las características relevantes de las HRTFs.

³⁶ La ITD correspondiente es restituida durante el procesamiento en tiempo real.

A modo de ejemplo, en la Figura 7.14 se muestra la comparación de la magnitud espectral de una HRTF medida y una estimada, para la posición $(\theta, \phi) = (0^\circ, -45^\circ)$ del oído derecho de un sujeto.

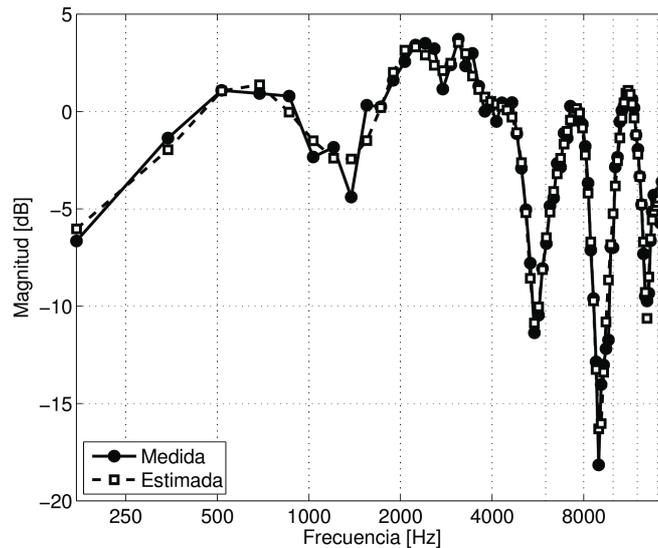


Figura 7.14 Magnitud espectral para una HRTF medida y estimada, para la posición 0° de azimut y -45° de elevación

7.5.3 Errores en el plano horizontal

En la Figura 7.15 se muestra la media del índice de correlación entre las HRIRs medidas y las estimadas de los 35 sujetos (línea negra) para el oído derecho en el plano horizontal. Los puntos indican los índices individuales de cada sujeto.

Se observa que en -55° , éstos son menores pero nunca disminuyen por debajo de 0,9. Cuando se producen los menores valores del índice, la cabeza oculta la ubicación de la fuente sonora al oído analizado (oído contralateral).

Morimoto (2001) investigó la importancia del oído contralateral e ipsilateral, para localizar una fuente sonora en el espacio. Concluyó que ambos oídos contribuyen de igual manera pero que la importancia del oído contralateral

disminuye a medida que la fuente se aleja lateralmente del plano medio, sobre el plano horizontal. Cuando este desplazamiento es mayor a 60° , la influencia del oído contralateral es irrelevante. Como los menores índices de correlación se producen en -55° , muy próximo al valor determinado por Morimoto, son perceptualmente irrelevantes.

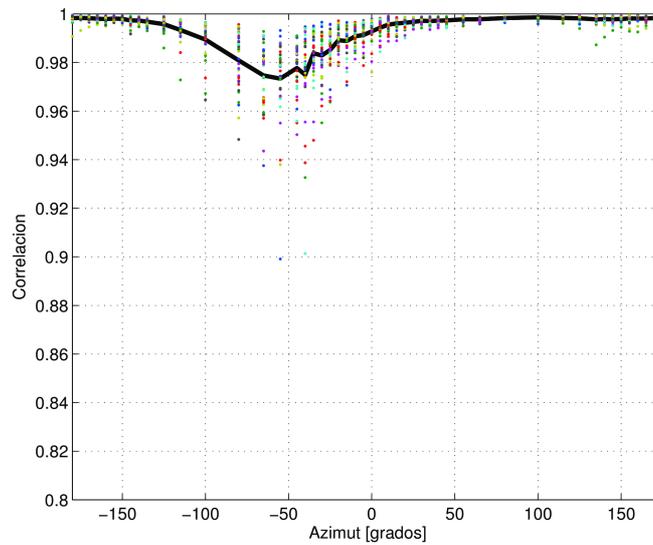


Figura 7.15 Correlación cruzada entre las HRIR medidas y las modeladas del oído derecho para el plano horizontal

7.5.4 Errores en las ILDs

La ILD comienza a ser importante para frecuencias donde la longitud de onda es menor al diámetro de la cabeza (sobre los 500 Hz aproximadamente). Sin embargo, se hace dominante para frecuencias por encima de los 1500 Hz. La ILD es mayor para ángulos de azimut de $+90^\circ$ y -90° (lateral izquierdo y derecho) y menor a 0° (justo al frente) y 180° (justo detrás).

La ILD para cada sujeto fue calculada mediante la siguiente fórmula

$$\text{ILD} = 20 \log \sqrt{\frac{1}{N} \sum_{i=1}^N H_r^2(i)} - 20 \log \sqrt{\frac{1}{N} \sum_{i=1}^N H_l^2(i)} \quad [\text{dB}] \quad (7.1)$$

donde N es la cantidad de componentes en frecuencia; H_r es la HRTF del oído derecho y H_l es la HRTF del oído izquierdo.

En la Figura 7.16 se muestra el error promedio de la ILD de los 35 sujetos en el plano horizontal para HRTFs medidas y estimadas en el rango de 1500 Hz a 15000 Hz. En puntos se muestran los errores individuales de los 35 sujetos. El error se estimó como la diferencia entre las ILDs de las HRTFs medidas y HRTFs estimadas. La mayoría de los errores individuales son menores a la jnd de 0,8 dB reportada por Hershkowitz y Durlach (1969).

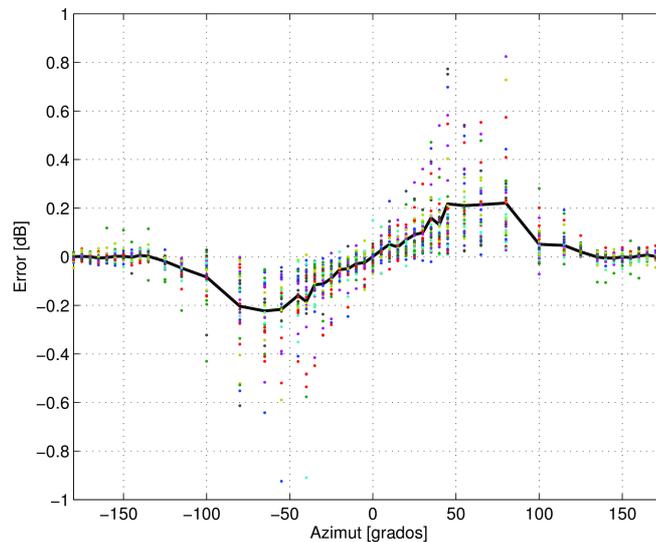


Figura 7.16 Error promedio de la diferencia entre las ILDs para HRTFs medidas y las ILD para HRTFs estimadas

7.5.5 Errores en las ITDs

Para analizar el comportamiento de las ITDs se utilizó la siguiente fórmula, dada por Kulkarni et al. (1999)

$$\text{ITD} = \tau(f) = \frac{\arg(H_r) - \arg(H_l)}{2\pi f} \quad (7.2)$$

donde \arg es una función que permite conocer la fase de la HRTF H , el subíndice hace referencia al oído derecho r y al izquierdo l , y f es la frecuencia. Esta fórmula fue aplicada tanto a las HRTFs medidas como a las estimadas.

Según Kulkarni et al., la ITD es una clave dominante hasta los 1500 Hz. En la Figura 7.17 se muestra el error promedio de la ITD de los 35 sujetos en el plano horizontal para HRTFs medidas y estimadas entre 172 Hz y 1500 Hz. En puntos se muestran los errores individuales de los 35 sujetos.

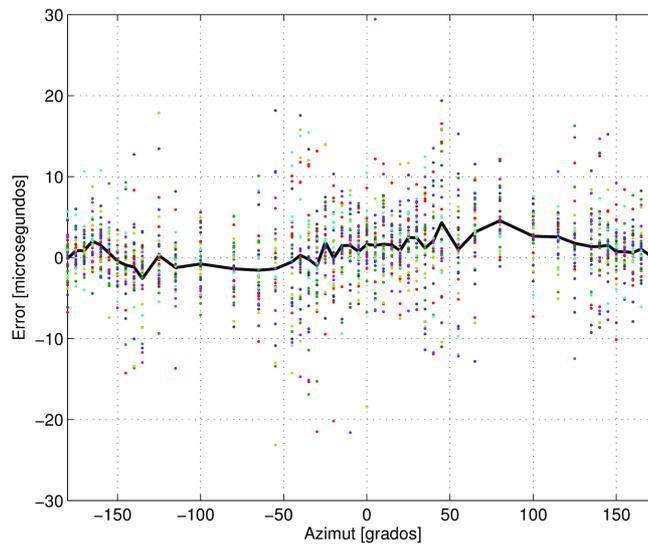


Figura 7.17 Error medio entre los ITD medidos y los ITD aproximados (trazo negro). En puntos, los errores individuales.

La mayoría de los errores son menores a la jnd de $20 \mu\text{s}$ reportada por Durlach

y Colburn (1978) (muy pocos puntos superan este valor). Kulkarni et al. (1999) establecieron que los sujetos no son sensibles a la fase monoaural, esto es, a la fase de los estímulos sonoros aplicados a cada oído. Demostraron, por el contrario, que el ser humano es altamente sensible a la diferencia de fase entre ambos oídos. Se comprobó que el comportamiento de los errores, para un sujeto determinado, no es sistemático en todo el plano horizontal, sino para posiciones determinadas (Ramos, Araneda, y Tommasini 2009). Esto podría ser consecuencia de fallas en las mediciones de las HRTFs (por ejemplo, movimientos involuntarios de la cabeza del sujeto).

7.5.6 Errores en el plano medio

Como se dijo, los contenidos espectrales de las HRTFs son dominantes en aquellas regiones del espacio en que las diferencias interaurales son nulas o ambiguas. Una de estas regiones es el plano medio. La interacción de las ondas sonoras con las cavidades y protuberancias de las orejas dan origen a resonancias (picos) y mínimos (notchs) que varían sistemáticamente con la posición de la fuente en este plano. Las HRTFs del oído derecho y del izquierdo son muy similares (difieren sólo debido a la asimetría natural del sujeto) y por lo tanto las claves están codificadas en los contenidos espectrales de las ondas que llegan a cada oído individualmente (claves monoaurales). En la Figura 7.18 se muestran las HRTFs de los 35 sujetos para una ubicación de la fuente sonora.

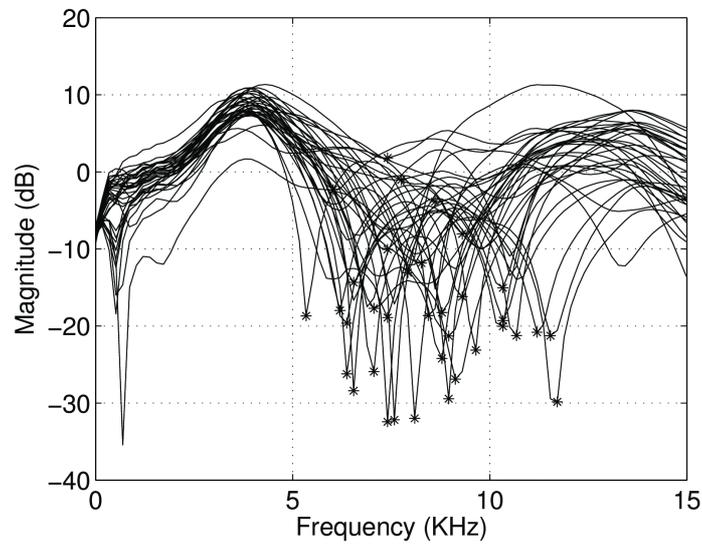


Figura 7.18 HRTFs de 35 sujetos

En dicha figura se observan resonancias y notchs bien definidos, producidos por interacción de las ondas sonoras en las orejas. Es oportuno resaltar las diferencias intersujeto, tanto en las frecuencias centrales como en la presión sonora.

Al igual que en el plano horizontal, se realizó la correlación cruzada entre las HRIRs medidas y las HRIRs obtenidas de los filtros (estimadas) (Figura 7.19). En trazo negro se muestran los índices de correlación promedio de los 35 sujetos correspondientes al oído derecho para el plano medio. Nuevamente, los puntos indican los índices individuales.

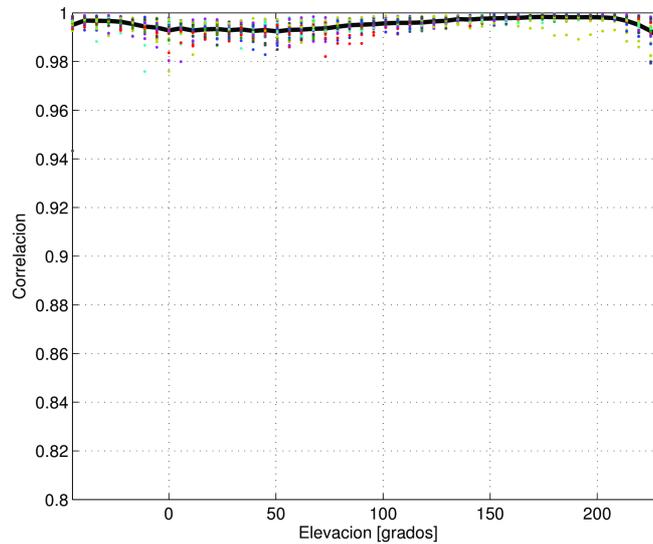


Figura 7.19 Correlación cruzada entre las HRIRs medidas y las HRIRs estimadas

El valor promedio del índice disminuye levemente desde -45° hasta 0° , se mantiene casi constante hasta los $+50^\circ$ y luego crece casi monótonicamente hasta los 180° , para volver a decrecer hasta los 230° . En otras palabras, los valores son menores por debajo del eje interaural y por encima hasta los $+50^\circ$. Cabe resaltar la escasa dispersión de los índices individuales para los 35 sujetos.

Moore et al. (1989) han investigado la detección y discriminación de un único notch, tanto de la frecuencia central como de la presión sonora, superpuesto a un ruido de banda ancha. Sus experimentos sugieren que los notchs en la región de frecuencias de 8 kHz son detectables, pero los cambios en la presión sonora de tales notchs son difíciles de discriminar. Los mismos autores afirmaron que errores de las frecuencias centrales de los notchs inferiores al 8,3% no se pueden discriminar. De igual manera establecieron que un ser humano puede discriminar diferencias de presión sonora mayores a 2-3 dB para presiones sonoras entre 3 y 9 dB.

En un estudio más reciente, Iida et al. (2007) se propusieron dilucidar cuáles picos y cuáles notchs, por encima de los 4 kHz, juegan un rol preponderante en el espectro de las HRTFs. En primer lugar, concluyeron que los cambios de las frecuencias centrales y de presión sonora de los notchs son claves importantes en la percepción de la elevación. Este hallazgo está de acuerdo con lo reportado por Hebrank y Wright (1974). Pero, en segundo lugar, indicaron que un único pico o un único notch no da información suficiente para juzgar la elevación de un sonido de banda ancha. El pico entre 4 y 5 kHz (modo 1 omnidireccional de Shaw) no colabora por sí solo en la percepción de la elevación. El sistema auditivo podría utilizar esta resonancia como una información de referencia para analizar los notchs. Sin embargo, últimos hallazgos han determinado que, si bien la frecuencia central del modo 1 es independiente de la posición, la presión sonora alcanza un máximo de 10 dB en las proximidades de 0° y disminuye por arriba y debajo de esta posición (Kahana y Nelson 2006; Ramos, Tommasini, y Araneda 2010). Esta variación puede ser discriminada (Moore, Oldfield, y Dooley 1989) y podría funcionar como una clave adicional.

Teniendo en cuenta las evidencias perceptuales mencionadas, se realizaron análisis complementarios en este plano para comprobar el grado de ajuste de las frecuencias centrales y de presión sonora, tanto de las resonancias como de los notchs característicos importantes entre -45° y +45°³⁷.

Primero se aplicó una ventana Hann modificada a ambas HRIRs, para aislar los efectos de las orejas (Algazi, Avendano, y Duda 2001) y eliminar las componentes de baja frecuencia que se producen por reflexiones en torso y hombros. En la Figura 7.20 se muestra un ejemplo de la ventana aplicada.

³⁷ Observar que en este rango los índices de correlación son levemente menores.

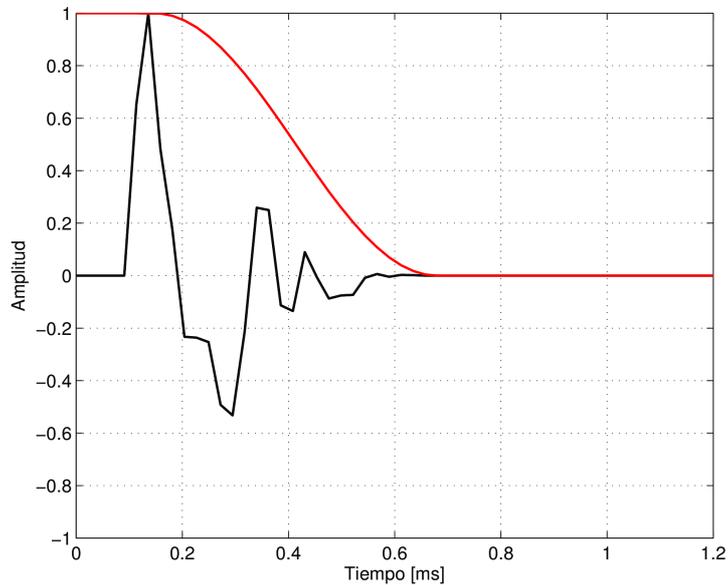


Figura 7.20 Ventana Hann modificada aplicada a las HRIRs

A continuación se midió, a un sujeto de la base de datos³⁸, las frecuencias centrales y los valores de presión sonora tanto para el modo 1 (4,2 kHz) como para el primer notch N1 (entre los 5 kHz y los 12 kHz). Este notch varía sistemáticamente con la posición de la fuente sonora entre -45° y $+45^\circ$ y es clave para percibir la ubicación de la fuente en el plano medio.

Para estimar los errores de las frecuencias centrales se utilizó la siguiente fórmula

$$E_f = \frac{f_c - \hat{f}_c}{f_c} \times 100 \quad [\%], \quad (7.3)$$

donde f_c es la frecuencia central de las HRTFs medidas; y \hat{f}_c es la frecuencia central de las HRTFs estimadas. Errores negativos significan que las frecuencias

38 Sujeto 003 de la base de datos de HRIRs del CIPIC. Utilizado para la simulación en este caso de estudio.

centrales de los notchs de las HRTFs estimadas, están desplazadas hacia las altas frecuencias, mientras que errores positivos significa lo contrario. Los errores de presión sonora se calcularon simplemente como la diferencia en decibeles.

Errores en el modo 1 de Shaw

En primer lugar se analiza el modo 1 de Shaw. En la Figura 7.21 se muestran las diferencias de presión sonora para la resonancia mencionada. Los errores positivos significan que los niveles de presión sonora de las HRTFs medidas son escasamente mayores que las del modelo.

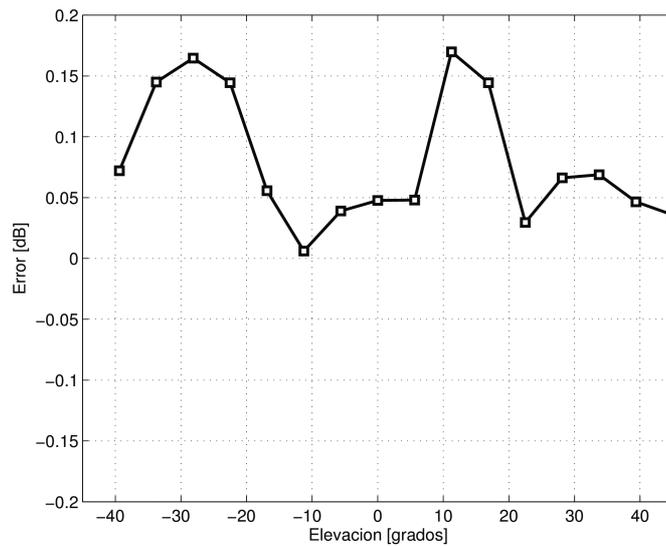


Figura 7.21 Error en el modo 1 de Shaw

Se comprueba que estas diferencias están muy por debajo del umbral de discriminación ($< 2-3$ dB). No se muestra los errores de las frecuencias centrales porque son nulos en el rango de posiciones analizadas.

Errores en el notch N1

En la Figura 7.22 se muestran los errores de presión sonora y de frecuencias centrales del primer notch. En ninguno de los dos casos se observa un comportamiento sistemático de los errores. Se ha comprobado que su distribución es totalmente diferente con otros sujetos de la muestra.

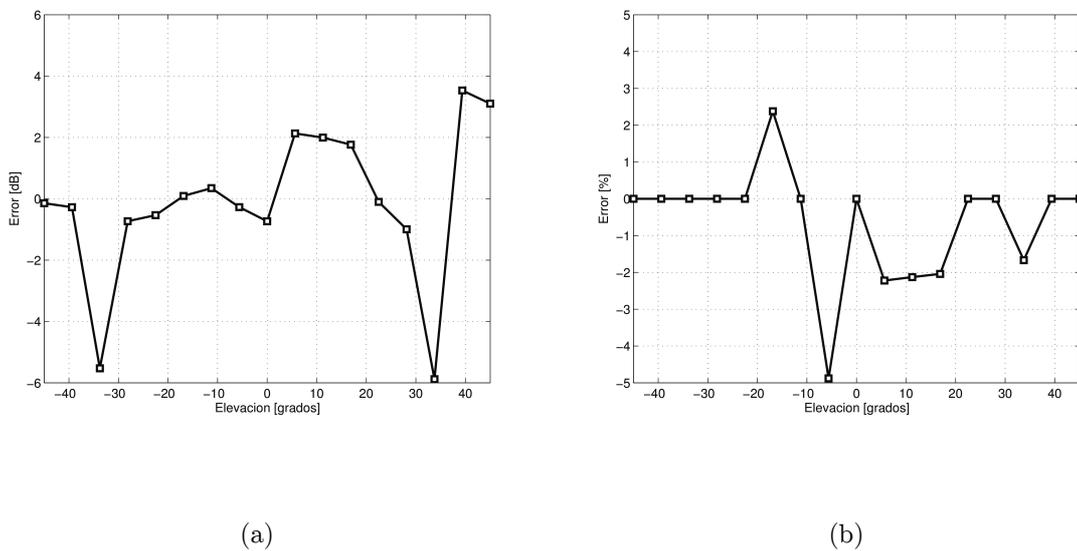


Figura 7.22 Errores de (a) presión sonora, y de (b) frecuencias centrales del primer notch

Analizando la Figura 7.22a se comprueba que los errores de presión sonora están dentro de lo estimado por Moore et al. (1989), salvo picos en dos posiciones de la fuente sonora ($-33,75^\circ$ y $+33,75^\circ$). Observando los errores de frecuencia, se verifica que, en todo el rango analizado, los errores están por debajo del umbral del 8,3%.

7.6 Validación del sensor de movimientos de cabeza

Para verificar si los valores de posición y orientación que arroja el sensor están dentro de los límites perceptuales requeridos por el AVRS, se realizaron mediciones de la precisión en el plano horizontal y en el plano medio.

7.6.1 Errores en el plano horizontal

En la Figura 7.23 se muestra el montaje experimental utilizado como referencia para la rotación azimutal en la medición. Consta de un disco de acrílico con orificios a lo largo de su borde exterior separados de manera muy precisa cada 10° . Al beacon se lo desmontó de los auriculares y se lo acopló al disco de manera solidaria. Cabe aclarar que todas las piezas que componen el sistema de referencia para el plano horizontal fueron mecanizadas mediante máquinas herramientas de control numérico de alta precisión. Lo mismo se hizo para el sistema de referencia del plano medio que se menciona más adelante (sección 7.6.2).



Figura 7.23 Sistema de referencia para el plano horizontal

Luego se hizo girar el disco en pasos de 10° hasta completar 360° (-170° a $+180^\circ$). Con ayuda de un emisor y un receptor infrarrojo (ubicado en la parte superior derecha de la figura) se logró determinar con exactitud la orientación del disco en el plano horizontal. Si se realizaba un acoplamiento óptico a través de uno de los orificios del disco, un circuito electrónico encendía un indicador lumínico. Así se corroboraba que la posición había variado exactamente 10° .

Se midieron 35 posiciones diferentes (360° espaciadas cada 10° , excluyendo a 0° que se tomó como referencia) en azimut, 10 veces cada una. En la Figura 7.24 se muestran los valores medios de las mediciones realizadas en el plano horizontal, junto con los valores del sistema de referencia. No existen diferencias apreciables entre la línea completa y la línea punteada.

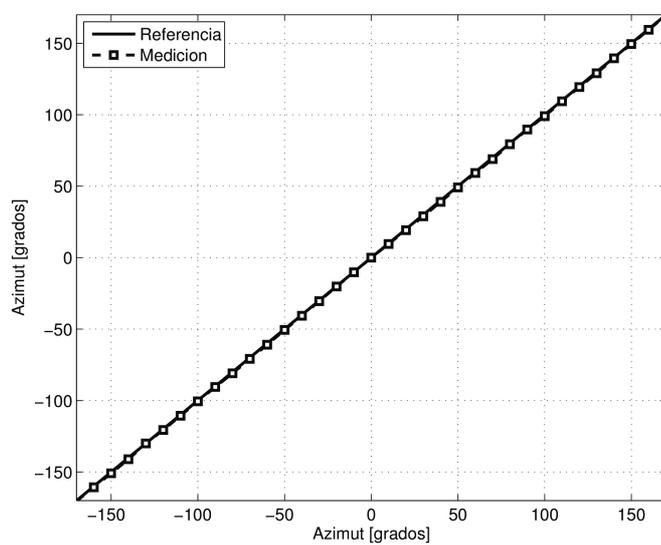


Figura 7.24 Valores de referencia y valores medidos en el plano horizontal.

El error medio fue de $0,62^\circ$ y la desviación estándar de $0,45^\circ$. El error máximo

ocurrió para un azimut de -140° y fue de $1,05^\circ$.

Como se dijo en el Capítulo 2, el umbral MAA (límite perceptual de discriminación de una fuente sonora) en el plano horizontal, para estímulos de banda ancha y para tonos de baja frecuencia, es de $1-2^\circ$ para las posiciones frontales (Mills 1958; Blauert 1997), se incrementa a $8-10^\circ$ en posiciones laterales extremas ($\pm 90^\circ$) (Kuhn 1987), y decrece nuevamente a $6-7^\circ$ si la fuente se ubica detrás del oyente (Mills 1958; Perrott 1969), siendo mayor para medias y altas frecuencias (Mills 1958; 1960; 1972). Para verificar si el sensor cumple estos umbrales, se realizó una prueba *t*-Student de una cola, con un nivel de significación $\alpha = 0,01$. Dado que los errores medidos siguen una distribución normal, se plantean las siguientes hipótesis (utilizando como umbral el promedio de los valores dados para la posición frontal, es decir, $1,5^\circ$)

$$\begin{aligned} H_0 : \mu_{\text{error}} &\geq 1,5^\circ \\ H_1 : \mu_{\text{error}} &< 1,5^\circ \end{aligned} \tag{7.4}$$

donde μ_{error} es la media del error.

H_0 fue rechazada ($t = -19,8034$, $p < 0,01$), por lo que la media del error es significativamente menor al umbral MAA para el plano horizontal en las posiciones frontales.

7.6.2 Errores en el plano medio

En el plano medio se midió el error variando la elevación del beacon en pasos de 5° desde una elevación de -45° hasta $+45^\circ$. Para lograr una medición confiable, el beacon se ubicó sobre un dispositivo de referencia que consta de un medio círculo que incluye una escala de precisión graduada en pasos de 5° (Figura 7.25).

Se midieron 18 posiciones diferentes (nuevamente se excluye a 0° que se toma como referencia) en elevación, 10 veces cada una. En la Figura 7.26 se muestran

los valores medios de las mediciones para el plano medio, junto con los valores del sistema de referencia.

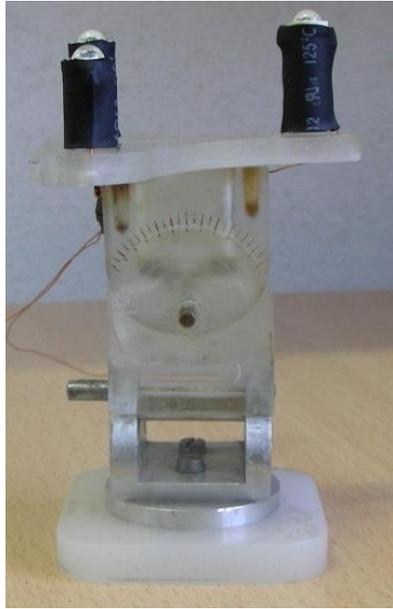


Figura 7.25 Sistema de referencia para el plano medio

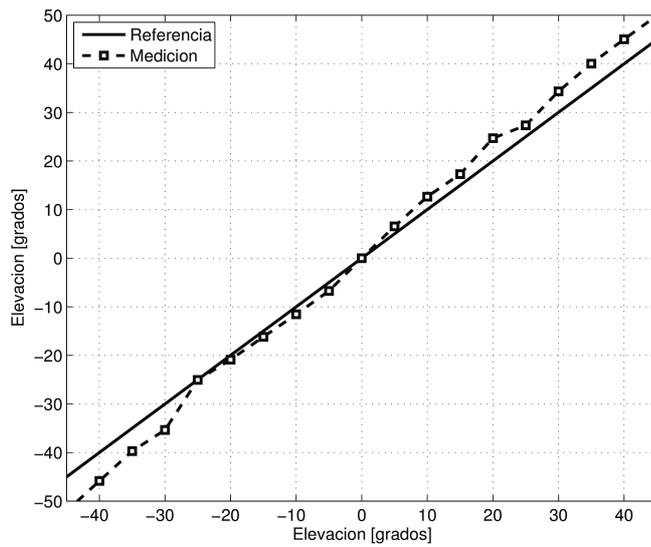


Figura 7.26 Valores de referencia y valores medidos en el plano medio.

El error medio fue de $3,34^\circ$ y la desviación estándar de $2,08^\circ$. El error máximo ocurrió para un valor de elevación de -45° y fue de $6,9^\circ$.

Como se dijo, el umbral MAA en el plano medio es de $3-9^\circ$ para fuentes ubicadas en el hemisferio frontal (Perrott y Saberi 1990; Blauert 1997). Nuevamente, para verificar si se cumplen estos límites, se realizó una prueba similar a la ecuación (7.1), con un nivel de significación $\alpha = 0,01$. Se plantearon las siguientes hipótesis (utilizando como umbral el promedio de los valores dados, es decir, 6°)

$$\begin{aligned} H_0 : \mu_{\text{error}} &\geq 6^\circ \\ H_1 : \mu_{\text{error}} &< 6^\circ \end{aligned} \quad (7.5)$$

Otra vez, H_0 fue rechazada ($t = -7,9119$, $p < 0,01$), por lo que la media del error fue significativamente menor al umbral MAA para el plano medio en el hemisferio frontal.

Se puede observar que para elevaciones mayores a $+30^\circ$ y menores -30° aproximadamente el error aumenta. Esto se debe a que las proyecciones de los puntos detectados en el plano de la cámara infrarroja se acercan entre sí a medida que aumenta la inclinación del beacon, provocando una pérdida de resolución. Puede llegar a ocurrir que la cámara del wiimote no pueda discriminar un punto infrarrojo del otro (ángulos mayores a $+50^\circ$, por ejemplo). Aunque el error actual es aceptable, se podría disminuir aún más con una redistribución de los puntos infrarrojos en el beacon.

7.7 Rendimiento del AVRS

El tiempo necesario para calcular la auralización depende de muchos factores, sin embargo los más influyentes son la geometría del recinto, la posición de la fuente sonora, la posición del oyente, la cantidad de VSs (calculadas y audibles), y

el orden de los filtros en la síntesis binaural.

Para este caso de estudio los tiempos que tomó calcular cada elemento de la cadena de filtros, con la configuración detallada en el Capítulo 6, se muestran en la Tabla 7.5.

	Tiempo [μ s]	Observaciones
Directividad	0,85	
Absorción de materiales	8,21	por superficie
Difusión	0,00	se realiza mediante FDN
Absorción del medio y distancia	0,52	
HRTF	89,31	para ambos oídos
Acumulación en el buffer	8,70	

Tabla 7.5 Tiempos de cálculo

Entonces, el tiempo que tardó el sistema para sintetizar a una reflexión especular de segundo orden (desde la fuente hasta el oyente) es de 0,115 ms. Para completar la parte de las reflexiones tempranas de la BIR (28 fuentes virtuales audibles de hasta segundo orden) se logró un tiempo de cálculo de 3,45 ms. La auralización completa del sonido tomó un tiempo de alrededor de 9 ms, para una BIR de 2,5 s (\approx 110000 muestras a 44,1 kHz).

7.8 Medición de los tiempos de latencia

Un parámetro crítico en el diseño de sistemas de simulación acústica virtual es el máximo tiempo de latencia que se puede permitir (umbral detectable) entre el movimiento de la cabeza del oyente y la actualización correspondiente del sonido auralizado.

Hay que tener en cuenta que la latencia total del sistema (total system latency,

TSL) o latencia de “extremo a extremo” y la frecuencia de actualización efectiva son parámetros distintos pero relacionados. La TSL está dada por las diferentes latencias de cada etapa individual del AVRS y la frecuencia de actualización del componente más lento determina la frecuencia de actualización efectiva.

Tanto la latencia como la frecuencia de actualización del sistema no se mantienen constantes a lo largo del tiempo. Por esta razón es usual reportar la media de un conjunto de mediciones, la desviación estándar y su rango.

En el AVRS, la TSL se puede dividir en varias etapas: el sensor de movimientos de cabeza, la auralización y la reproducción del sonido (Figura 7.27). Aunque también se deben tener en cuenta los costos de las comunicaciones y de intercambio de datos (que se incluyeron en cada etapa). Se realizaron mediciones de latencia de los eventos indicados en la figura: cambio de posición u orientación de la cabeza, recepción de los datos del movimiento y reproducción de los datos auralizados.

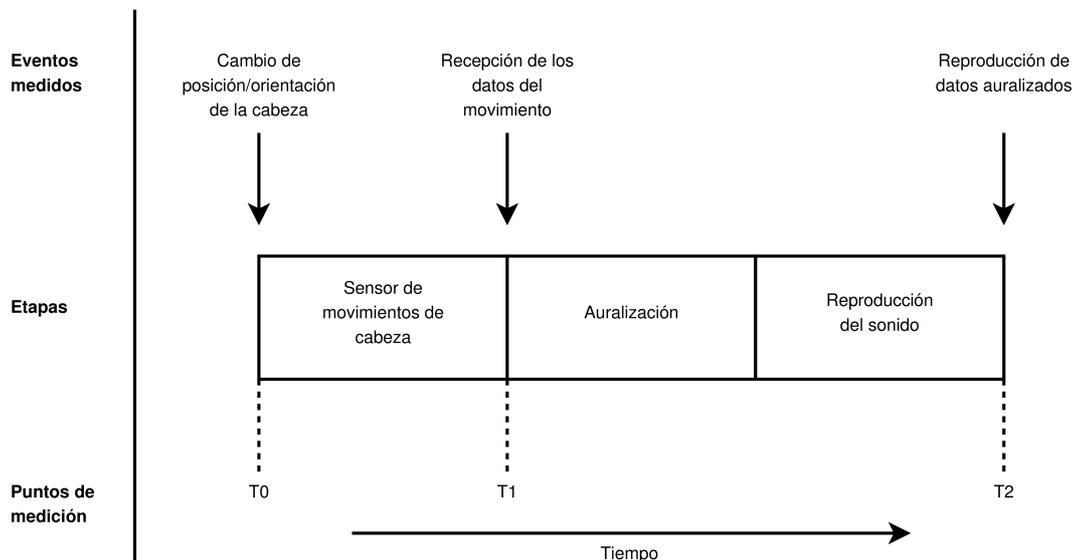


Figura 7.27 Eventos y etapas para medir la TSL

7.8.1 Latencia del sensor de movimientos de cabeza

En la Figura 7.28 se muestra un diagrama de la configuración experimental para la medición de la latencia del sensor de movimientos. Notar que el beacon fue removido de los auriculares y se lo montó nuevamente sobre el disco de acrílico utilizado anteriormente (Figura 7.23). De esta forma se puede lograr un patrón de movimiento que es repetible, facilitando la medición.

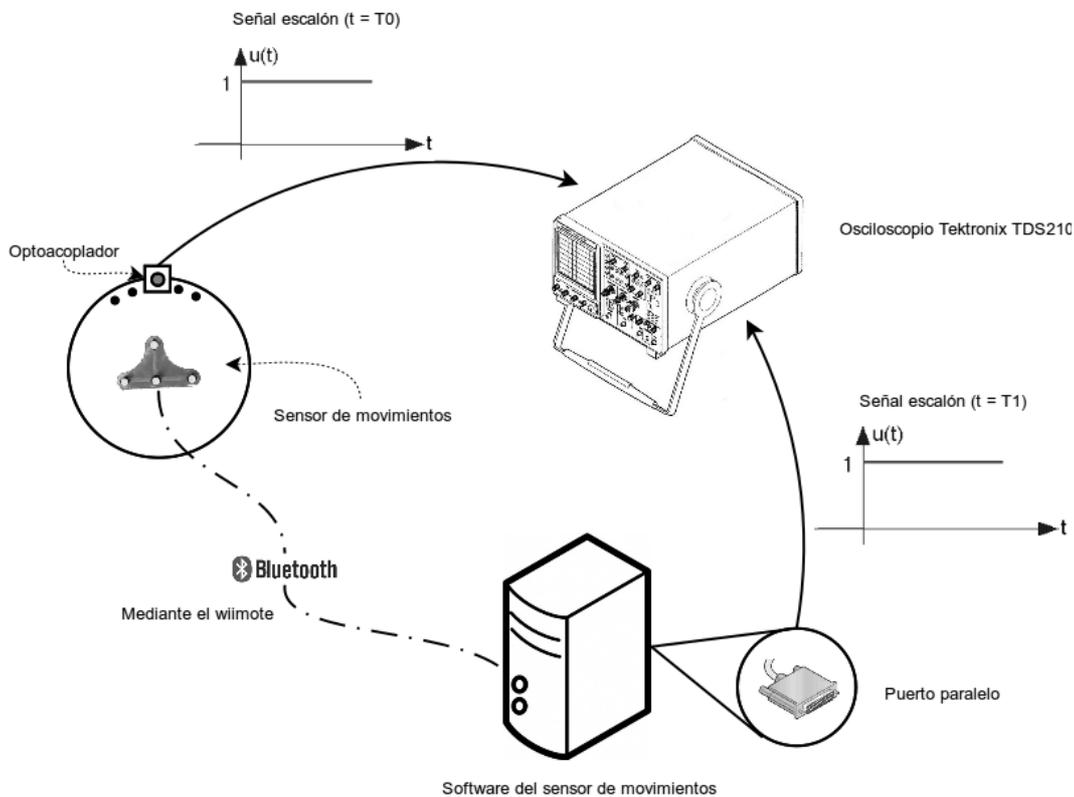


Figura 7.28 Configuración experimental para la medición de la latencia

Se eligió uno de los orificios del disco de acrílico como punto de referencia (ángulo de azimut 0°) y el resto de los orificios fueron obstruidos. Este valor de azimut fue usado por el software para comparar con los datos que posteriormente fueron enviados por el sensor. Cuando el disco pasa por el punto de referencia, el

optoacoplador emite una señal escalón (tiempo T0, ver Figura 7.27). Ahora, cuando el software detecta que la posición dada por el sensor coincide con la de referencia, se genera otra señal escalón que es enviada a través de un *pin* del puerto paralelo³⁹ (tiempo T1). La latencia del puerto se puede despreciar ya que está en el orden de algunos microsegundos.

Tanto la señal generada por el optoacoplador como la señal del puerto paralelo fueron grabadas en un osciloscopio digital (Tektronix TDS210). La diferencia del tiempo T1 con el tiempo T0 es la latencia del sensor de movimientos de cabeza.

En la Tabla 7.6 se resumen los resultados de 30 mediciones.

	Latencia [ms]
Media	26,89
Desviación estándar	6,09
Mínimo	17,00
Máximo	38,40

Tabla 7.6 Valores de latencia del sensor de movimientos de cabeza

Dependiendo de su valor, la latencia del sensor de movimientos de cabeza puede tener dos tipos de efectos sobre el rendimiento de la simulación acústica virtual (Brungart, Simpson, y Kordik 2005).

El primero está relacionado con errores por parte del oyente al juzgar la ubicación de una fuente sonora, que se pone en evidencia por: una disminución en la precisión y/o un incremento en el tiempo de localización.

Hay diferentes trabajos que estudian la latencia máxima de un sensor en tareas que requieren precisión para localizar una fuente sonora. En general, esta latencia

³⁹ Para enviar la señal se utilizó la librería open-source Parapin para GNU/Linux (parapin.sourceforge.net).

depende de la duración del estímulo que se reproduce. Para estímulos de larga duración (8 s), la localización es relativamente precisa incluso con latencias tan grandes como 500 ms (Wenzel 1999), a excepción de algunas posiciones en el plano medio. Estudios con estímulos de menor duración (1,5 a 3 s), han mostrado que recién se observa un deterioro en la localización con latencias de aproximadamente 96 ms (Sandvad 1996; Wenzel 2001). Para estímulos breves (125 ms a 2 s), latencias mayores a 73 ms tienden a disminuir la precisión en la localización e incrementar el tiempo de respuesta (Brungart et al. 2004).

El segundo efecto perjudicial que tiene la latencia del sensor de movimientos es la “pérdida de realismo” (y el aumento de la fatiga) que ocurre cuando el retardo es tal que el oyente lo puede percibir. No existen muchos antecedentes que indiquen cuál es el mínimo retardo que el ser humano es capaz de detectar.

Brungart et al. (2005) realizaron experimentos en este sentido. Mostraron que un oyente promedio es incapaz de detectar latencias menores a 80 ms en un sistema con una única fuente virtual. Los sujetos que obtuvieron mejor rendimiento en las pruebas fueron incapaces de detectar latencias menores a 60 ms. Esto sugiere que las latencias del sensor comienzan a ser notables en aproximadamente el mismo punto donde empieza a afectar el rendimiento en tareas de localización para estímulos breves. Según toda esta evidencia, la latencia del sensor de movimientos se encuentra por debajo de los umbrales perceptuales requeridos, aún para estímulos breves.

Lentz (2007) capturó los movimientos de cabeza de un usuario inmerso en un sistema de realidad virtual (VirKopf/RAVEN) mientras interactuaba con el entorno. Luego calculó la velocidad angular promedio de los movimientos de cabeza, que es de $8,686^{\circ}/s$, con un valor máximo de $141,458^{\circ}/s$. Esta información

se usó para calcular el error promedio del sensor de movimientos debido a su latencia. El error ronda los 0,23°, con un error máximo de 3,8°.

7.8.2 Latencia total del AVRS

El montaje para la medición de la TSL (que incluye la latencia del sensor de movimientos) fue similar al utilizado en la medición anterior. Pero ahora, en lugar de emplear el puerto paralelo para enviar la señal, se utilizó la placa de sonido. Cuando la posición dada por el sensor coincide con la de referencia, el AVRS actualiza el escenario acústico virtual realizando los cálculos de la etapa de auralización (ver Figura 7.27). Luego, en lugar de enviar a la placa de audio el sonido auralizado, envió una onda cuadrada⁴⁰ (tiempo T2). De esta manera se incluyó en la medición el proceso de cálculo y la latencia inherente a la placa de sonido (Intel HDA con codec Realtek ALC888). Haciendo la diferencia de ambos tiempos (T2 y T0) con el osciloscopio, se calculó la TSL.

En la Tabla 7.7 se resumen los resultados.

	TSL [ms]
Media	46,13
Desviación estándar	17,09
Mínimo	23,00
Máximo	77,00

Tabla 7.7 Valores de latencia extremo a extremo

No hay muchos estudios que indaguen sobre la mínima latencia detectable en un sistema de realidad acústica virtual. Wenzel (1998) dedujo esta latencia partir de resultados de pruebas psicoacústicas sobre el mínimo ángulo audible de una

⁴⁰ Dado que no es posible reproducir una salida de corriente continua en la placa de sonido.

fuelle en movimiento (minimum audible movement angle, MAMA) (Perrott 1977). 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). Sin embargo, un estudio reciente (Lindau 2009), no encontró relación entre la velocidad máxima del movimiento de cabeza y dicho umbral de la TSL.

En los últimos años, algunos autores (Mackensen 2004; Yairi, Iwaya, y Suzuki 2006; Lindau 2009) midieron la mínima latencia detectable en diferentes entornos acústicos virtuales (AVEs). Los umbrales promedio reportados varían desde 60 a 85 ms, con un valor mínimo en 53 ms obtenido por Lindau. Estos valores indican que la TSL del AVRS está por debajo de los umbrales reportados.

8 – Conclusiones

El objetivo de la presente tesis fue desarrollar un sistema experimental dinámico e interactivo de realidad acústica virtual (el AVRS), basado en arquitectura de computadoras de propósito general y que sea flexible, escalable, fácil de mantener y modular. Para ello se aplicaron modelos físico-matemáticos simplificados, basados en evidencias perceptuales de la fuente, el recinto y el medio de propagación, y el oyente, aptos para ser ejecutados en tiempo real y reproducidos mediante auriculares.. Se diseñó e implementó un sistema de código abierto programado bajo el paradigma orientado a objetos y se evaluó la verosimilitud de los modelos mediante un caso de estudio.

8.1 Síntesis de la implementación y validación del sistema

A continuación se sintetiza la implementación de cada subsistema (fuente, recinto y medio, y oyente) y del sistema completo, junto a la validación de los mismos. Además, se describen los aportes realizados que permitieron cumplir con los objetivos propuestos.

Fuente sonora

El modelado de la fuente sonora (Capítulo 3) es una parte fundamental en los

sistemas de realidad acústica virtual. En el AVRS se utilizó el método de filtrado direccional. Se lograron buenos resultados con filtros IIR de quinto orden, en la misma dirección que Savioja et al. (1999). Esto representa una ventaja con respecto a otros proyectos de simulación acústica de código abierto, como son slab3d (Wenzel, Miller, y Abel 2000), RTVAS (Scarpaci 2006) y SSR (Ahrens, Geier, y Spors 2008), entre otros, en los cuales la fuente sonora no está implementada.

Recinto y medio de propagación

El recinto es otro de los componentes esenciales del escenario acústico (Capítulo 4). La respuesta impulsiva de un recinto (RIR) se divide en tres partes: sonido directo, reflexiones tempranas (especulares y difusas) y reflexiones tardías.

El sonido directo y las reflexiones tempranas son relevantes para que un ser humano localice con precisión una fuente sonora y experimente una verídica percepción del espacio en el que está inmerso. El sonido directo y las reflexiones tempranas especulares fueron modelados mediante un método exacto: el de la fuente-imagen (ISM). Se desarrolló un nuevo algoritmo para la prueba de audibilidad de fuentes virtuales de orden superior (Tommasini y Ramos 2008) y se redujo el costo computacional respecto a lo propuesto por Borish (1984).

Para modelar las reflexiones tempranas difusas, inicialmente se propusieron dos métodos: a) filtros diseñados a partir de modificaciones a los propuestos por Pellegrini (2002) (Tommasini, Ramos, y Ferreyra 2008a), y b) el uso de la primera parte de la respuesta de una red de retardo realimentada (FDN). Se optó por la FDN debido a su mejor rendimiento, ya que no insume recursos porque la respuesta es calculada “off-line”.

Si bien las reflexiones tardías son irrelevantes para localizar una fuente, permiten al oyente percibir otras características del recinto, tales como dimensiones, calidad acústica y distancia a la fuente sonora. La FDN también fue utilizada para generar las reflexiones tardías debido a que facilita el control sobre ellas y garantiza los aspectos perceptuales asociados a la densidad de tiempo y de frecuencia (Tommasini, Ramos, y Ferreyra 2008a).

Oyente

Para modelar al oyente (Capítulo 5) se aproximaron las respuestas impulsivas de cabeza (HRIRs) o funciones de transferencia de cabeza (HRTFs), mediante filtros IIR, diseñados a partir del método iterativo de Steiglitz-McBride (Ramos, Araneda, y Tommasini 2009a). Se consiguió un buen ajuste con filtros de orden 20 y con cinco iteraciones y se evaluaron objetivamente teniendo en cuenta aspectos perceptuales (ver más adelante). Se consiguió una economía en el uso de la memoria y una disminución en el tiempo de ejecución.

En el mundo real, el movimiento relativo entre una fuente sonora y un oyente es percibido como un fenómeno auditivo continuo. Una persona adulta puede discriminar un cambio de posición de una fuente ubicada justo al frente en el plano horizontal, de tan sólo 1° a 2° y entre 3° y 9° en el plano medio y hemisferio frontal. No se dispone de HRTFs medidas con esa resolución espacial. Langendijk y Bronkhorst (2000) determinaron que, aplicando interpolación bilineal a HRTFs medidas con una resolución espacial de alrededor 6° , es posible obtener HRTFs intermedias sin que se deteriore la precisión para localizar fuentes sonoras. En esta tesis se utilizó interpolación bilineal para obtener HRTFs con una resolución espacial de $1,25^\circ$ en el plano horizontal y de $2,8125^\circ$ en el plano medio (Tommasini

y Ramos 2010). La HRTF de una posición no medida, se obtuvo mediante la combinación lineal de las cuatro HRTFs más cercanas, pesadas por coeficientes que dependen de la distancia entre la posición interpolada y las medidas.

Cabe recordar que, para que un sujeto experimente una auténtica percepción del espacio, es necesario disponer de sus propias HRTFs. Como estas mediciones son complejas, costosas y requieren de equipamiento especial, es importante el desarrollo de métodos para personalizar HRTFs medidas a otras personas (HRTFs “genéricas”). Se han realizado avances en este aspecto. Se construyeron HRTFs personalizadas a partir de la estimación de los pesos por regresión lineal múltiple (Ramos, Araneda, y Tommasini 2009b) y por redes neuronales artificiales (ANNs), utilizando mediciones antropométricas de las orejas. Se utilizó la distorsión espectral (SD), índice ampliamente utilizado en la literatura, para realizar las comparaciones entre ambos métodos. Los resultados obtenidos en el plano medio indicaron que las ANNs son mejores que la regresión lineal múltiple. Sin embargo se pudo comprobar que la SD no es un índice apropiado para evaluar la calidad de ajuste en el plano medio⁴¹. Se pudo comprobar que los valores de los errores entre las HRTFs medidas y las HRTFs estimadas son detectables y discriminables por el ser humano y podrían producir consecuencias perceptuales adversas. La utilización de medidas unidimensionales de las cavidades y protuberancias de la oreja, podría ser adecuado para estimar algunas de las resonancias características del oído externo. Para determinar las frecuencias centrales y las presiones sonoras de los notchs característicos, entre -45° y $+45^\circ$ ⁴², es esencial que se reporte las formas de la oreja, en especial, la de la pared posterior de la concha.

41 Este índice oculta las diferencias espectrales por su carácter integrador respecto de la frecuencia.

42 Rango donde los notchs son importantes.

AVRS

La implementación del sistema de simulación acústica virtual en tiempo real se desarrolló en el Capítulo 6. Para concretarlo, en primer lugar, se realizó una serie de programas basados en scripts de MATLAB. Los cuales fueron: un simulador de acústica de recintos, con el cual fue posible sintetizar la RIR y visualizar la trayectoria de los rayos sonoros (Tommasini y Ramos 2008); un módulo para calcular los parámetros acústicos especificados en la norma ISO 3382 (2009) y evaluar las RIRs simuladas (Tommasini, Ramos, y Ferreyra 2008b); y un programa que contempló diversos modelos del oyente para evaluarlos tanto objetivamente como subjetivamente, mediante pruebas psicoacústicas (Ramos, Araneda, y Tommasini 2009a; Ramos, Calvo, y Tommasini 2007).

En segundo lugar, se migraron los modelos ya validados al sistema en tiempo real. El AVRS se ejecuta sobre un sistema operativo de tiempo real (RTOS) GNU/Linux con RTAI-LXRT, lo que aseguró cumplir con las restricciones críticas de tiempo (ver más adelante) (Tommasini y Ramos 2010). La complejidad y precisión de los modelos están limitados solamente por los recursos de hardware disponibles.

El AVRS se compone de: una interfaz de configuración, un módulo de procesos a nivel de usuario y un módulo de procesos en tiempo real (tareas *hard real-time*). Los procesos en tiempo real tienen un período de ejecución de aproximadamente 11,6 ms (512 muestras a 44,1 kHz). En ese tiempo se actualiza la posición y orientación del oyente, se realizan los cálculos de los modelos para obtener la respuesta impulsiva binaural (BIR), se convoluciona la BIR con una señal anecoica de entrada y se reproduce la señal ya procesada (auralizada). Cada uno de estos

subprocesos se ejecutan en paralelo.

Las reflexiones tempranas fueron procesadas por una cadena de filtros. Estos filtros tuvieron en cuenta: directividad de la fuente sonora, características acústicas de los materiales, atenuación por la distancia y las HRTFs. Es de suma importancia que los coeficientes de absorción utilizados para la creación de los filtros de materiales sean fiables, ya que se comprobó que una mínima diferencia en sus valores pueden provocar errores significativos en la simulación (Tommasini, Ramos, y Ferreyra 2008b).

Las reflexiones tardías se generaron con una FDN de 8 líneas. Esta red permite ajustar el tiempo de reverberación dependiente de la frecuencia, la densidad temporal y la densidad de frecuencia de las reflexiones (Tommasini, Ramos, y Ferreyra 2008a).

Las reflexiones tempranas (junto al sonido directo) y las reflexiones tardías deben ser empalmadas de manera que la transición sea continua e imperceptible al oyente. Basándose en análisis estadísticos de la reverberación, Jot et al. (1997) establecieron que el tiempo a partir del cual la RIR se asemeja a un proceso estocástico está dado por un parámetro conocido como mixing time. Uniendo ambos tipos de reflexiones en este tiempo, se obtuvo la BIR (Tommasini y Ramos 2008).

Finalmente, se realizó la convolución de baja latencia entre la BIR y un sonido anecoico para producir la señal de salida (auralizada). Esta convolución sólo agrega un retardo de 11,6 ms independiente del tamaño de la BIR.

Para intensificar la vivencia del oyente en el entorno virtual y asemejar dicha sensación a una situación real, es crucial incorporar al AVRS la captura de los movimientos de cabeza. Para ello se diseñó e implementó un sensor de

movimientos de cabeza de un costo relativamente bajo, basado en hardware estándar y software libre. Los valores medios de los errores tanto en el plano horizontal como en el plano medio (0.62° y 3.14° respectivamente) fueron aceptables, ya que son significativamente menores a los umbrales reportados por la literatura ($1-2^\circ$ y $3-9^\circ$ respectivamente) (Mills 1958; Blauert 1997). Como se vio, esto también está en concordancia con la resolución espacial lograda para las HRTFs.

Evaluación y validación

Los criterios de evaluación del sistema se basaron en las mínimas diferencias que un ser humano es capaz de detectar y discriminar. Estas diferencias se conocen como las diferencias apenas notables o mínimas diferencias notables (just-noticeable difference, jnd) y sus valores se desprenden de pruebas psicoacústicas.

Para validar el sistema globalmente, se simuló la presencia de un sujeto en una aula de clases de la Universidad Tecnológica Nacional, Facultad Regional Córdoba (Capítulo 7). Las propiedades acústicas del aula fueron evaluadas mediante la determinación de algunos de los parámetros aconsejados en la norma ISO 3382 que se calculan a partir de la RIR. Estos índices están relacionados con aspectos subjetivos, referentes a las sensaciones que un individuo experimenta cuando está presente en un recinto en particular. El método de evaluación consistió en comparar los parámetros obtenidos de la respuesta impulsiva medida del aula (RIR medida) y de la RIR simulada. Se verificó que, para los tiempos de reverberación, las diferencias o errores están por debajo de las jnd establecidas. Sólo unos pocos valores de los parámetros de energía superan, para algunas frecuencias, las jnd. No se han encontrado en la literatura otros trabajos que validen los resultados de la

simulación de recintos en tiempo real utilizando evidencias perceptuales.

Como se mencionó, para modelar al oyente se diseñaron filtros IIR, a partir de las HRTFs medidas a un grupo de personas. Las características relevantes de las ondas sonoras que llegan a ambos tímpanos de un oyente son: la diferencia de nivel y de tiempo entre ambos oídos (ILD e ITD respectivamente) y los perfiles espectrales. Estas características varían sistemáticamente con la ubicación de la fuente sonora y son utilizadas por el sistema auditivo del oyente, para juzgar dicha posición. La habilidad y precisión del ser humano en esta tarea es sorprendente y, en consecuencia, es fundamental reproducir estas claves auditivas. Se compararon las HRTFs medidas con las producidas por los filtros y se utilizó como criterio de aceptabilidad las jnd respectivas (Ramos, Araneda, y Tommasini 2009a).

En el plano horizontal se observó que los valores de correlación entre las HRTFs medidas y estimadas fueron muy altos, cercanos a 1. Los valores menores del índice (entre 0,9 y 0,95) se produjeron para el oído contralateral, en posiciones perceptualmente irrelevantes.

Los errores en ILDs fueron menores a 0,5 dB para la mayoría de los sujetos. Los casos en los cuales estuvieron próximos a 1 dB se presentaron para posiciones del oído contralaterales a la fuente.

Los errores en ITDs fueron, en su mayoría, menores a la jnd (20 μ s). Se comprobó que el comportamiento de los errores para un determinado sujeto, no fue sistemático en todo el plano horizontal, sino para posiciones aisladas. Esto podría ser consecuencia de fallas en la medición de las HRTFs (por ejemplo, movimientos involuntarios de la cabeza del sujeto).

En el plano medio, se obtuvo una correlación casi perfecta y hubo escasa dispersión en los índices de correlación individuales. Teniendo en cuenta evidencias

perceptuales señaladas por varios autores (Moore, Oldfield, y Dooley 1989; Iida et al. 2007), se realizaron análisis complementarios en este plano, para comprobar el grado de ajuste de las frecuencias centrales y de presión sonora tanto para las resonancias como para los notchs característicos entre -45° y $+45^\circ$.

Se comprobó que los errores de presión sonora en la resonancia característica (modo 1 de Shaw) estuvieron muy por debajo del umbral de discriminación (2-3 dB) y los errores de las frecuencias centrales fueron despreciables.

Para el notch característico (N1), también se comprobó que los errores de presión sonora estuvieron, en general, por debajo del umbral de discriminación. Respecto a los errores de las frecuencias centrales, se verificó que se encontraron por debajo de la jnd (8,3 %) en todo el rango analizado.

La función de transferencia total desde la fuente hasta los tímpanos del oyente debe ser actualizada cuando éste se mueve. Esto significa que se deben cambiar los coeficientes de los filtros de toda la cadena. Se evaluó el rendimiento del AVRS en términos de tiempos de cálculo y de latencia. Los resultados arrojaron que el tiempo que tarda el sistema para procesar una reflexión especular de segundo orden (desde la fuente sonora al oyente) es de 0,115 ms. La construcción de una BIR de 2,5 s de largo (≈ 110000 muestras a 44,1 kHz) le insumió al AVRS un tiempo de aproximadamente 9 ms. Es oportuno resaltar aquí que, como el sistema es escalable, estos tiempos podrían mejorarse con un hardware más rápido.

Por último, se realizaron mediciones de los tiempos de latencia del sensor de movimientos y del sistema completo (desde que el oyente se mueve hasta que le llega el sonido auralizado). La latencia del sensor de movimientos de cabeza puede tener dos tipos de efectos perceptuales. El primero está relacionado con errores por parte del oyente al juzgar la ubicación de una fuente sonora. El segundo es la

“pérdida de realismo” que ocurre cuando el oyente percibe un retardo entre lo que escucha y su propio movimiento. El valor de latencia promedio del sensor (26,89 ms), se encuentra por debajo de los umbrales especificados (entre 73 y 80 ms). La velocidad angular promedio de los movimientos de cabeza, reportados por Lentz (2007), es de $8,686^\circ/\text{s}$. Teniendo esto en cuenta, se calculó el error promedio del sensor de movimientos debido a su latencia. Este error fue $0,23^\circ$, que sumado al error de precisión propio del sensor, está por debajo de las jnd, tanto en el plano horizontal ($1\text{-}2^\circ$ en posiciones frontales) como en el plano medio ($3\text{-}9^\circ$ en el hemisferio frontal).

En relación a la latencia total del sistema (TSL), cabe destacar que, recientemente algunos autores midieron la mínima latencia detectable en entornos acústicos virtuales (Mackensen 2004; Yairi, Iwaya, y Suzuki 2006; Lindau 2009). Estos trabajos reportan umbrales que varían entre 60 y 85 ms. El valor promedio de la TSL del sistema desarrollado en la presente tesis es de 46,13 ms. Como se comprueba, está por debajo de los umbrales mencionados.

8.2 Principales aportes y conclusión general

Los principales aportes de la tesis se pueden resumir en los siguientes puntos:

- Se desarrolló un sistema de realidad acústica virtual original dinámico e interactivo, en el cual se ha modelado los siguientes subsistemas: la fuente sonora (mediante filtros IIR de directividad); el recinto y el medio de transmisión (síntesis de la RIR mediante una metodología híbrida); y el oyente (mediante filtros IIR para las HRTFs). Se desarrolló un sensor de posición de cabeza de bajo costo que brinda la realimentación necesaria al usuario. No hay antecedentes de trabajos que simulen un entorno auditivo

virtual con este abordaje.

- El AVRS es un sistema open-source desarrollado bajo un sistema operativo en tiempo real (RTOS) GNU/Linux con RTAI-LXRT y utilizando hardware estándar. Logra cumplir todas las restricciones críticas de tiempo: la latencia total del sistema es de aproximadamente 46 ms, por debajo de los umbrales perceptuales reportados en la literatura.
- Se evaluaron objetivamente los modelos, adoptando como criterio de aceptabilidad las diferencias apenas notables (jnd) pertinentes. Los resultados obtenidos mostraron ser aceptables y, en general, están por debajo de las jnd. Tampoco se han encontrado trabajos que validen la simulación en tiempo real utilizando este criterio.
- Se hicieron avances para personalizar HRTFs en el plano medio, mediante redes neuronales y regresión lineal múltiple. Se demostró que la distorsión espectral (SD) no es un estimador adecuado del error en este plano, ya que no tiene en cuenta las diferencias espectrales. Se debe profundizar sobre cuáles son las mediciones antropométricas del torso, hombros y orejas que influyen en las HRTFs, para diseñar estrategias que permitan construir HRTFs de cualquier individuo sin tener que medirlas.

En conclusión, se cumplen todos los requerimientos exigidos a un sistema de realidad acústica virtual. El AVRS sintetiza sonidos espacializados de manera eficiente, robusta y con baja latencia.

Para finalizar, es importante destacar que este sistema es una genuina plataforma experimental flexible que por un lado facilitará la investigación de aspectos aún no resueltos de la simulación acústica y, por el otro, permitirá su

utilización como un verdadero banco de pruebas en nuevas aplicaciones no tradicionales.

8.3 Trabajo a futuro

El AVRS desarrollado en esta tesis, por ser un sistema abierto y flexible, permite extender y optimizar sus actuales capacidades. A continuación se detallan algunos aspectos que se tiene previsto realizar en una próxima etapa del trabajo.

- Los recintos de geometría compleja, donde la cantidad de superficies es elevada y sus formas son irregulares, demandan un poder de cálculo mayor. Es común asumir en el modelado acústico de los mismos, que las superficies tienen formas convexas. Pero algunos recintos tienen superficies con ángulos reentrantes (la forma de la superficie deja de ser convexa), lo que puede provocar una obstrucción en el camino de un sonido. Tener en cuenta estas evidencias, originan una carga extra al procesamiento de las reflexiones tempranas, pero amplía las posibilidades del modelado de recintos.
- Una prestación importante de un sistema de realidad acústica virtual, es la posibilidad de manejar múltiples fuentes sonoras. De esta manera, se podría representar una orquesta, donde cada integrante sea una fuente sonora independiente (con su posición, orientación y características de directividad). También es útil cuando se pretende considerar fuentes no correlacionadas, tales como, el ruido de un tren, lluvia en el techo o el aplauso del público.
- En el presente sistema, tanto la difusión como la reverberación se modelaron de manera aproximada. Estas simplificaciones fueron aceptables desde el punto de vista perceptual y lograron muy buena eficiencia computacional.

Pero, con el objetivo de ampliar la aplicación del AVRS al campo de la acústica arquitectónica, sería deseable contar con una simulación más exacta desde el punto de vista físico-acústico. El método de trazado de rayos (ray-tracing) es una alternativa posible. Para ello se deberán implementar nuevos métodos y tecnologías que permitan acelerar el tiempo de cálculo de esta técnica.

- La incorporación de nuevos modelos y funcionalidades y el agregado de complejidad a los algoritmos existentes hace que la auralización en tiempo real requiera de mayor poder de cómputo. Esto amerita considerar nuevas estrategias de implementación, tales como la computación de alto rendimiento. Una buena opción son los procesadores gráficos (graphic processing units, GPUs) de las placas de video. En los últimos años, la arquitectura de las GPUs ha evolucionado drásticamente sobrepasando el rendimiento de los procesadores de propósito general (central processing units, CPUs). Este alto rendimiento es debido a la paralelización masiva de tareas. Algunos investigadores han comenzado a utilizar estas GPUs programables en aplicaciones de realidad acústica virtual (Wefers y Schröder 2009; Savioja, Manocha, y Lin 2010; Tsingos, Jiang, y Williams 2011). Se tiene previsto adoptar esta tecnología para resolver los puntos mencionados más arriba.
- Durante el desarrollo de la tesis se hicieron pruebas informales con sujetos para validar la sensación de inmersión que tiene un oyente al utilizar el AVRS. Pero es necesario validar al sistema mediante la implementación de pruebas psicoacústicas. Se pretende desarrollar estas pruebas con la participación de un equipo interdisciplinario de profesionales, que incluye a

ingenieros, psicólogos y fonoaudiólogos.

8.4 Publicaciones relacionadas a la tesis

Las publicaciones vinculadas al desarrollo de esta tesis son las siguientes (listadas en orden alfabético por apellido de los autores):

- Arias, C., Hüg, M. X., Ramos, O. A., **Tommasini, F. C.**, Novillo, D., y Esquinas, P. 2006. Ecolocación Humana: Movimientos de cabeza para localizar sonidos. *V Congreso Iberoamericano de Acústica (FIA 2006)*, paper ID: A006. Santiago de Chile, Chile.
- Arias, C., Ramos, O. A., Ortiz Skarp, A. H., Hüg, M. X., Gómez, C., Bermejo, F., **Tommasini, F. C.**, Esquinas, P., y Barrera, F. 2007. Movimientos de cabeza para localizar sonidos reflejados. En M. S. Ison y M. C. Richaud (Eds.), *Avances en Investigación en Ciencias del Comportamiento en Argentina* (1ra ed.): 903–920. Mendoza: Editorial Universidad del Aconcagua. ISBN 978-987-23232-5-7.
- Cravero, G. A., Ferreyra, S., Ramos, O. A., **Tommasini, F. C.**, y Araneda, M. 2011. Base de datos de mediciones de respuestas impulsivas de recintos con diferentes características acústicas. *Mecánica Computacional*, XXX: 3125-3142. ISSN 1666-6070.
- Ramos, O. A., Araneda, M., y **Tommasini, F. C.** 2008. Análisis de patrones de las funciones de transferencia de cabeza. *Actas del VI Congreso Iberoamericano de Acústica (FIA 2008)*, paper ID: A092. Buenos Aires, Argentina. ISBN 978-987-24713-1-6.
- Ramos, O. A., Araneda, M., y **Tommasini, F. C.** 2009. Construcción de HRTF individuales a partir de mediciones antropométricas. *Mecánica*

- Computacional*, XXVIII: 149–161. ISSN 1666-6070.
- Ramos, O. A., Araneda, M., y **Tommasini, F. C.** 2009. Diseño y evaluación de filtros binaurales. *Mecánica Computacional*, XXVIII: 137-148. ISSN 1666-6070.
 - Ramos, O. A., Arias, C., **Tommasini, F. C.**, Ortiz Skarp, A. H., Novillo, D., Bermejo, F., Hüg, M. X, y Lunati, V. 2009. Movimientos de cabeza en una prueba de ecolocación. *Mecánica Computacional*, XXVIII: 123-135. ISSN 1666-6070.
 - Ramos, O. A., Calvo, G., y **Tommasini, F. C.** 2007. Modelo acústico de cabeza y torso mediante análisis de componentes principales. *Mecánica Computacional*, XXVI: 46–58. ISSN 1666-6070.
 - Ramos, O. A., **Tommasini, F. C.**, y Araneda, M. 2010. Contribución de la cabeza, el torso y el oído externo en las funciones de transferencia relativas a la cabeza. *2º Congreso Internacional de Acústica UNTREF*, Buenos Aires, Argentina.
 - Ramos, O. A., y **Tommasini, F. C.** 2008. Simplificación de las funciones de transferencia de cabeza mediante análisis de componentes principales. *Mecánica Computacional*, XXVII: 431–442. ISSN 1666-6070.
 - **Tommasini, F. C.**, Ramos, O. A., y Ferreyra, S. 2008. Comparación objetiva de respuestas impulsivas de recintos medidas y simuladas. *Actas del VI Congreso Iberoamericano de Acústica (FIA 2008)*, paper ID: A087. Buenos Aires, Argentina. ISBN: 978-987-24713-1-6.
 - **Tommasini, F. C.**, Ramos, O. A., y Ferreyra, S. 2008. Modelado eficiente y validación objetiva de la respuesta impulsiva de un recinto. *Mecánica Computacional*, XXVII: 443-455. ISSN 1666-6070.

- **Tommasini, F. C.**, y Ramos, O. A. 2008. Modelo basado en evidencias perceptuales para simular respuestas impulsivas de recintos. *Anales de la XXXIV Conferencia Latinoamericana de Informática (CLEI 2008)*: 1100-1109. Santa Fe, Argentina. ISBN 978-950-9770-02-7.
- **Tommasini, F. C.**, y Ramos, O. A. 2010. Auralización en tiempo real: implementación del modelo del oyente. *2º Congreso Internacional de Acústica UNTREF*, Buenos Aires, Argentina.

Apéndices

A – Parámetros acústicos de recintos

La norma ISO 3382 (2009) define un conjunto de parámetros acústicos para caracterizar el campo sonoro de un recinto. Aquí se detallan los utilizados en esta tesis.

A.1 Parámetros relacionados a la reverberación

El tiempo de reverberación es una de las mediciones más importantes que se realizan en un recinto. W. C. Sabine fue el primero en introducir el concepto de tiempo de reverberación. Lo definió como el intervalo de tiempo (en segundos) que se requiere para que la energía sonora decaiga 60 dB.

Sin embargo, en muy pocas ocasiones es posible medirlo en forma directa, debido al elevado nivel del ruido de fondo. La norma ISO 3382, plantea diferentes métodos para medir este parámetro, siendo el método de la integración inversa de la RIR el empleado en esta tesis. Para su obtención, básicamente se eleva al cuadrado los valores instantáneos de presión sonora de la RIR y, seguidamente, se realiza una integración inversa (se suma la energía desde la derecha a la izquierda), lo que es conocido como la integración de Schroeder o *backward integration*. Así se

obtiene la curva del decaimiento de la energía sonora (Figura A.1), denominada *curva de decaimiento* (ISO 2009).

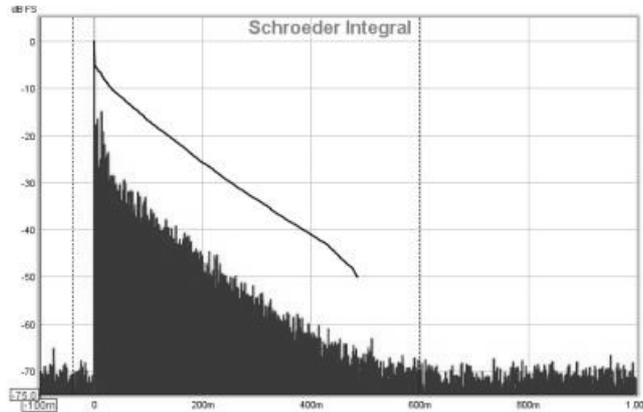


Figura A.1 Integración de Schroeder

En una situación ideal, sin ruido de fondo⁴³, el decaimiento como función del tiempo es

$$E(t) = \int_t^{\infty} p^2(\tau) d\tau = \int_{\infty}^t p^2(\tau) d(-\tau) \quad (\text{A.1})$$

donde p es la respuesta impulsiva.

Esta integral en tiempo inverso se resuelve a menudo de la siguiente manera

$$\int_t^{\infty} p^2(\tau) d\tau = \int_0^{\infty} p^2(\tau) d\tau - \int_0^t p^2(\tau) d\tau. \quad (\text{A.2})$$

Como la parte inicial del decaimiento es la más interesante, se utiliza la relación de caída (en un rango que varía según el parámetro) como estimador del tiempo de reverberación. Los parámetros EDT (early decay time), T_{10} , T_{20} y T_{30} se miden a partir de la pendiente de la curva de decaimiento. Esta pendiente se

⁴³ Para una situación real, existen varias técnicas para minimizar el ruido de fondo de la RIR (Hirata 1982; Lundeby et al. 1995; Morgan 1997; Faiget et al. 1998).

determina a partir de la línea que mejor ajusta a la curva en el rango apropiado, utilizando el método de regresión lineal. Luego, extrapolando la línea, se calcula el tiempo de reverberación como el tiempo necesario para decaer 60 dB⁴⁴.

Los rangos de ajuste para los diferentes parámetros son:

- EDT se basa en el ajuste de una línea entre 0 dB a -10 dB.
- T_{10} se basa en el ajuste de una línea entre -5 dB a -15 dB.
- T_{20} se basa en el ajuste de una línea entre -5 dB a -25 dB.
- T_{30} se basa en el ajuste de una línea entre -5 dB a -35 dB.

La norma ISO 3382 (2009) recomienda calcular tanto al EDT como los tiempos de reverberación (T_{10} , T_{20} y T_{30}), si se verifica un determinado nivel de relación señal-ruido (SNR). Jordan (1981) demuestra que el EDT se correlaciona de mejor manera con la percepción de la reverberación, mientras que los tiempos de reverberación están relacionados con las propiedades físicas del recinto.

A.2 Parámetros relacionados a la energía

Existen variados parámetros o índices que establecen relaciones entre la energía temprana y tardía contenida en la RIR. Han sido ampliamente utilizados en estos últimos años para evaluar la calidad acústica de un recinto en lo referente a la inteligibilidad de la palabra y a la claridad o definición de las producciones musicales.

El parámetro propuesto por Reichardt et al. (1975), actualmente especificado por la norma para condiciones relacionadas a la palabra, establece el límite de tiempo de la energía temprana en 50 ms. Este parámetro es denotado como C_{50} y su fórmula de cálculo es

⁴⁴ Se asume que la parte de la curva usada es representativa del decaimiento completo.

$$C_{50} = 10 \log_{10} \left(\frac{\int_0^{0,050 \text{ s}} p^2(t) dt}{\int_{0,050 \text{ s}}^{\infty} p^2(t) dt} \right) \quad [\text{dB}]. \quad (\text{A.3})$$

De manera similar, esta relación se puede aplicar para condiciones relacionadas a la música. Aquí el límite de tiempo de la energía temprana es de 80 ms, y el parámetro es C_{80} (usualmente llamado “claridad”). La fórmula es

$$C_{80} = 10 \log_{10} \left(\frac{\int_0^{0,080 \text{ s}} p^2(t) dt}{\int_{0,080 \text{ s}}^{\infty} p^2(t) dt} \right) \quad [\text{dB}]. \quad (\text{A.4})$$

También es posible medir la relación entre la energía temprana y la energía total. Este parámetro, nombrado D_{50} , fue introducido originalmente por Thiele (1953) y luego fue incluido en la norma ISO 3382. El acrónimo “D” proviene de la palabra “deutlichkeit” que, en alemán, significa “claridad”. Sin embargo fue traducida al inglés como “definition” y al español como “definición”. D_{50} es usualmente utilizado para condiciones relacionadas a la palabra y se calcula como

$$D_{50} = \frac{\int_0^{0,050 \text{ s}} p^2(t) dt}{\int_0^{\infty} p^2(t) dt} . \quad (\text{A.5})$$

Otra opción para este grupo de parámetros, es el propuesto por Kürer (1969), denominado “centro de tiempo” o T_s , el cual representa el tiempo del centro de gravedad energético de la respuesta impulsiva. Su fórmula es

$$T_s = \frac{\int_0^{\infty} tp^2(t) dt}{\int_0^{\infty} p^2(t) dt} \quad [\text{s}]. \quad (\text{A.6})$$

B – Sistemas de referencia de la cabeza

Hay diferentes sistemas de coordenadas que permiten especificar la posición de una fuente sonora respecto a la cabeza del oyente. En este apéndice se explican los utilizados en esta tesis.

B.1 Planos y ejes

Para especificar la ubicación de una fuente sonora, se definen tres planos característicos. El plano horizontal o azimutal contiene una línea imaginaria que une ambos oídos denominado eje interaural y define dos hemisferios: el inferior y el superior; el plano medio o vertical, que contiene al eje longitudinal y es perpendicular al eje interaural, define los hemisferios derecho e izquierdo; y el plano frontal, perpendicular a los dos anteriores, determina el hemisferio delante y detrás del sujeto (Figura B.1).

Al oído que se encuentra en el mismo hemisferio que la fuente respecto al plano medio, se lo denomina ipsilateral y al oído que queda oculto por la cabeza, contralateral.

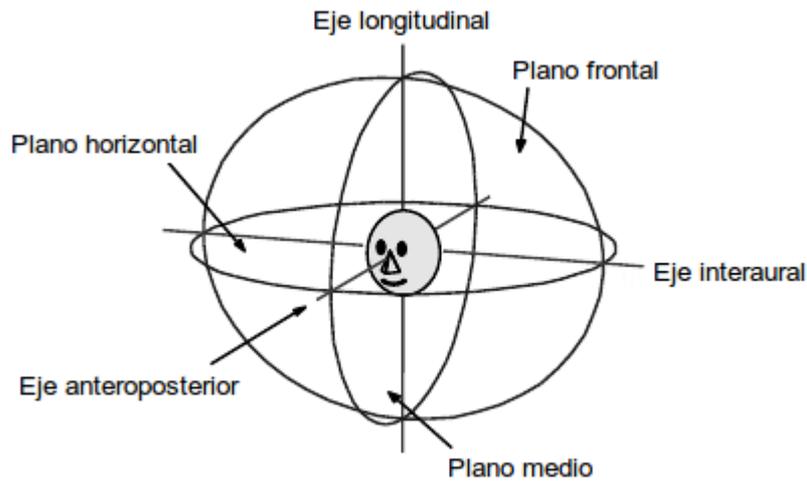


Figura B.1 Planos y ejes de referencia de la cabeza. Adaptado de Scarpaci (2006).

B.2 Sistemas de coordenadas

Dos tipos de sistemas de coordenadas se utilizan, en general, para especificar la posición de la fuente sonora respecto al oyente: coordenadas polares referidas al eje longitudinal y coordenadas polares referidas al eje interaural.

B.2.1 Coordenadas polares referidas al eje longitudinal

Este sistema de coordenadas es el más intuitivo, ya que imita el movimiento físico de la cabeza. Asumiendo que existe un vector que une al oyente con la fuente sonora, la posición de una fuente sonora se especifica por dos ángulos que se definen de la siguiente manera:

- **Azimut** (θ): es el ángulo entre el eje anteroposterior y el vector proyectado sobre el plano horizontal. Si $\theta = 0^\circ$ la fuente se encuentra ubicada sobre el plano medio. Si θ es negativo la fuente está a la izquierda del plano medio y si θ es positivo, a la derecha.
- **Elevación** (ϕ): es el ángulo entre el vector y el plano horizontal. Si $\phi = 0^\circ$

la fuente se encuentra ubicada sobre el plano horizontal. Si ϕ es negativo la fuente está por debajo del plano horizontal y si ϕ es positivo por arriba.

Los valores de θ están restringidos al intervalo $(-180^\circ, +180^\circ]$, y los valores de ϕ al intervalo $[-90^\circ, +90^\circ]$.

En la Figura B.2 se muestra un ejemplo de este sistema de coordenadas.

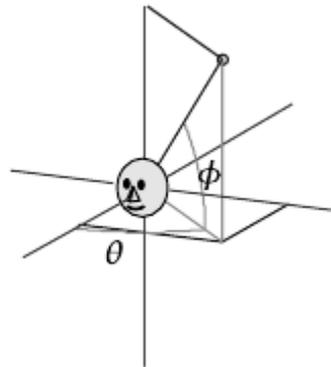


Figura B.2 Coordenadas polares referidas al eje longitudinal

B.2.2 Coordenadas polares referidas al eje interaural

En este sistema de coordenadas, el ángulo de azimut está íntimamente relacionado a la ITD, ya que define los conos de confusión independientemente de la elevación.

Nuevamente, asumiendo que existe un vector que une al oyente con la fuente sonora, la posición de una fuente sonora se especifica por los siguientes ángulos:

- **Azimut** (θ): es el ángulo entre la ubicación del vector y el plano medio. Si $\theta = 0^\circ$ la fuente se encuentra ubicada sobre el plano medio. Si θ es negativo la fuente está a la izquierda del plano medio y si θ es positivo, a la derecha.
- **Elevación** (ϕ): es el ángulo entre el eje anteroposterior y el vector

proyectado sobre el plano medio. Si $\phi = 0^\circ$ la fuente se encuentra ubicada sobre el plano horizontal. Si ϕ es negativo la fuente está por debajo del plano horizontal y si ϕ es positivo por arriba.

En la Figura B.3 se muestra un ejemplo de este sistema de coordenadas.

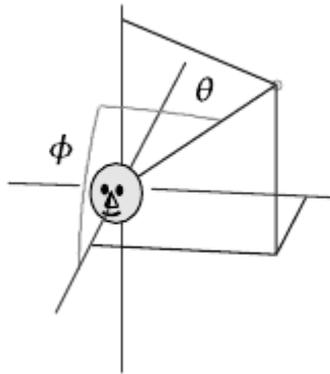


Figura B.3 Coordenadas referidas al eje interaural

Se podrá notar que cuando la fuente está en el plano horizontal o en el plano medio, ambos sistemas de coordenadas coinciden.

Referencias

- Abbott, D. 2006. *Linux for Embedded and Real-time Applications* (2nd ed.). Newnes.
- Ahearn, M. J., Schaeffler, M. J., Celmer, R. D. y Vigeant, M. C., 2009. The Just Noticeable Difference in the Clarity Index for Music, C80. *The Journal of the Acoustical Society of America*, 126 (4): 2288.
- Ahnert, W., y Feistel, R. 1992. EARS Auralization Software. *Preprints of the Audio Engineering Society 93th Convention*, 3388.
- Ahrens, J., Geier, M., y Spors, S. 2008. The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. *Preprints of the Audio Engineering Society 124th Convention*, 7330.
- Algazi, V. R., Avendano, C., y Duda, R. O. 2001. Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109 (3): 1110-1122.
- Algazi, V. R., Duda, R. O, Duraiswami, R., Gumerov, N. A., y Tang, Z. 2002. Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, 112 (5): 2053-2064.
- Algazi, V. R., Duda, R. O, Morrison, R. P., y Thompson, D. M. 2001. Structural composition and decomposition of HRTFs. *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*: 103–106, New Paltz, NY.
- Algazi, V. R., Duda, R. O, Thompson, D. M., y Avendano, C. 2001. The CIPIC

- HRTF database. *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*: 99–102. New Paltz, NY.
- Allen, J. B., y Berkley, D. A. 1979. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4): 943–950.
- Arias, C. 2009. *Ecolocación humana y efecto precedente*. Tesis doctoral. Facultad de Psicología, Universidad Nacional de Córdoba.
- Arias, C., Hüg, M. X., Ramos, O. A., Tommasini, F. C., Novillo, D., y Esquinas, P. 2006. Ecolocación Humana: Movimientos de Cabeza para Localizar Sonidos. *Actas del V Congreso Iberoamericano de Acústica*: 1–9, Santiago de Chile.
- Arias, C., Ramos, O. A., Ortiz Skarp, A. H., Hüg, M. X., Gómez, C., Bermejo, F., Tommasini, F. C., Esquinas, P., y Barrera, F. 2007. Movimientos de cabeza para localizar sonidos reflejados. En M. S. Ison y M. C. Richaud (Eds.), *Avances en Investigación en Ciencias del Comportamiento en Argentina* (1ra ed.): 903–920. Mendoza: Editorial Universidad del Aconcagua.
- Arias, C., y Ramos, O. A. 2003. Audición espacial en ambientes reverberantes: aspectos teóricos relevantes. *Interamerican Journal of Psychology*, 37(2): 373-382.
- Asano, F., Suzuki, Y., y Sone, T. 1990. Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, 88(1): 159-168.
- Ashmead, D. H., Davis, D. F., y Northington, A. 1995. Contribution of listeners' approaching motion to auditory distance perception. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2): 239–256.
- Ashmead, D. H., LeRoy, D., y Odom, R. D. 1990. Perception of the relative distances of nearby sound sources. *Perception & Psychophysics*, 47(4): 326-331.
- Barbalace, A., Luchetta, A., Manduchi, G., Moro, M., Soppelsa, A., y Taliercio, C. 2008. Performance comparison of VxWorks, Linux, RTAI, and Xenomai in

- a hard real-time application. *IEEE Transactions on Nuclear Science*, 55(1): 435–439.
- Bass, H. E., Bauer, H.-J., y Evans, L. B. 1972. Atmospheric Absorption of Sound: Analytical Expressions. *The Journal of the Acoustical Society of America*, 52(3B): 821-825.
- Basso, G. 2006. *Percepción auditiva*. Editorial Universidad Nacional de Quilmes, Buenos Aires, Argentina.
- Begault, D. R. 1994. *3-D sound for virtual reality and multimedia*. AP Professional, NY.
- Beranek, L. L. 1986. *Acoustics*. American Institute of Physics, NY.
- Beranek, L. L. 1996a. Acoustics and musical qualities. *The Journal of the Acoustical Society of America*, 99(5): 2647-2652.
- Beranek, L. L. 1996b. *Concert and Opera Halls: How They Sound* (1st ed.). Acoustical Society of America, Woodbury, NY, USA.
- Blauert, J. 1997. *Spatial hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, MA.
- Blauert, J., Lehnert, H., Sahrhage, J., y Strauss, H. 2000. An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation. *Acta Acustica united with Acustica*, 86(1): 94-102.
- Blessner, B. A. 2001. An Interdisciplinary Synthesis of Reverberation Viewpoints. *Journal of the Audio Engineering Society*, 49(10): 867–903.
- Blommer, M. A., y Wakefield, G. H. 1997. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Transactions on Speech and Audio Processing*, 5(3): 278-287.
- Borish, J. 1984. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6): 1827-1836.
- Bos, L., y Embrechts, J. J. 2009. An interactive real-time based auralization system for room acoustics implementing directional impulse response and multiple audio reproduction modules for spatialization (the AURALIAS

- project). *Proceedings of NAG/DAGA*: 1459–1462, Rotterdam.
- Botteldooren, D. 1995. Finite-difference time-domain simulation of low-frequency room acoustic problems. *The Journal of the Acoustical Society of America*, 98(6): 3302-3308.
- Bradley, J. S., y Soulodre, G. A. 1995. The influence of late arriving energy on spatial impression. *The Journal of the Acoustical Society of America*, 97(4): 2263-2271.
- Brown, C. P., y Duda, Richard O. 1998. A structural model for binaural sound synthesis. *IEEE Transactions on Speech and Audio Processing*, 6(5): 476-488.
- Brown, C., y May, B. 2005. Comparative Mammalian Sound Localization. En R. R. Fay y A. N. Popper (Eds.), *Sound Source Localization*: 124-178. Springer, NY.
- Brungart, D. S., Simpson, B. D., McKinley, R. L., Kordik, A. J., Dallman, R. C., y Ovenshire, D. A. 2004. The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. *Proceedings of International conference on Auditory Display (ICAD)*: 6–9.
- Brungart, D. S., Simpson, B. D., y Kordik, A. J. 2005. The detectability of headtracker latency in virtual audio displays. *Proceedings of International conference on Auditory Display (ICAD)*: 37–42.
- Bucher, R., y Balemi, S. 2005. Scilab/Scicos and Linux RTAI - a unified approach. *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005*: 1121–1126.
- Buttazzo, G. C. 2004. *Hard real-time computing systems: predictable scheduling algorithms and applications*. Springer-Verlag, NY.
- Carrión Isbert, A. 2001. *Diseño acústico de espacios arquitectónicos*. Alfaomega, Barcelona.
- Chen, J., Veen, B. D., y Hecox, K. E. 1995. A spatial feature extraction and regularization model for the head-related transfer function. *The Journal of the Acoustical Society of America*, 97(1): 439-452.

- Christensen, C. L., y Rindel, J. H. 2005a. Predicting acoustics in class rooms. *Proceedings of the 34th International Congress and Exposition on Noise Control Engineering (Inter-Noise 2005)*. Rio de Janeiro, Brazil.
- Christensen, C. L., y Rindel, J. H. 2005b. A new scattering method that combines roughness and diffraction effects. *Forum Acousticum*. Budapest, Hungary.
- Chu, D., y Stanton, T. K. 1990. Application of Twersky's boss scattering theory to laboratory measurements of sound scattered by a rough surface. *The Journal of the Acoustical Society of America*, 87(4): 1557-1568.
- Coleman, P. D. 1962. Failure to Localize the Source Distance of an Unfamiliar Sound. *The Journal of the Acoustical Society of America*, 34(3): 345-346.
- Cravero, G. A., Ferreyra, S., Ramos, O. A., Tommasini, F. C., y Araneda, M. 2011. Base de datos de mediciones de respuestas impulsivas de recintos con diferentes características acústicas. *Mecánica Computacional*, XXX: 3125-3142.
- Daigle, J. N., y Xiang, N. 2006. A specialized fast cross-correlation for acoustical measurements using coded sequences. *The Journal of the Acoustical Society of America*, 119(1): 330-335.
- Dalenbäck, B.-I. 1996. Room acoustic prediction based on a unified treatment of diffuse and specular reflection. *The Journal of the Acoustical Society of America*, 100(2): 899-909.
- Dalenbäck, B.-I., Kleiner, M., y Svensson, U. P. 1993. A Macroscopic View of Diffuse Reflection. *Preprints of the Audio Engineering Society 95th Convention*: 3726.
- Dementhon, D. F., y Davis, L. S. 1995. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15(1-2): 123-141.
- Djelani, T., Pörschmann, C., Sahrhage, J., y Blauert, J. 2000. An Interactive Virtual-Environment Generator for Psychoacoustic Research II: Collection of Head-Related Impulse Responses and Evaluation of Auditory Localization. *Acta Acustica united with Acustica*, 86(6): 1046-1053.
- Duda, R. O., y Martens, W. L. 1998. Range dependence of the response of a

- spherical head model. *The Journal of the Acoustical Society of America*, 104(5): 3048-3058.
- Durlach, N. I., y Colburn, H. S. 1978. Binaural phenomena. In E. C. Carterette y M. P. Friedman (Eds.), *Handbook of perception*: 365–466. Academic Press, NY.
- Eyring, C. F. 1930. Reverberation Time in “Dead” Rooms. *The Journal of the Acoustical Society of America*, 1(2A): 168.
- Faiget, L., Legros, C., y Ruiz, R. 1998. Optimization of the Impulse Response Length: Application to Noisy and Highly Reverberant Rooms. *Journal of the Audio Engineering Society*, 46(9): 741–750.
- Farina, A. 1995. RAMSETE - a new Pyramid Tracer for medium and large scale acoustic problems. *Proceedings of Euro-Noise 95*. Lyon, France.
- Farina, A. 2000. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Preprints of the Audio Engineering Society 108th Convention*. Paris, France.
- Ferreira, S., y Ramos, O. A. 2007. Análisis físico-acústico-espacial de respuestas impulsivas de un recinto con alta dispersión sonora obtenidas por métodos indirectos. *Mecánica Computacional*, XXVI: 3-14.
- Firestone, F. A. 1930. The phase difference and amplitude ratio at the ears due to a source of pure tone. *The Journal of the Acoustical Society of America*, 2(2): 260-270.
- Fleisch, T. 2009. Wiimote Virtual Reality Desktop. *Coding4Fun*. Recuperado el 16 de octubre de 2009 de:
<http://channel9.msdn.com/coding4fun/articles/Wiimote-Virtual-Reality-Desktop>.
- Frenette, J. 2000. *Reducing artificial reverberation algorithm requirements using time-variant feedback delay networks*. Master's thesis, University of Miami, Miami.
- Freyman, R. L., Clifton, R. K., y Litovsky, R. Y. 1991. Dynamic processes in the precedence effect. *The Journal of the Acoustical Society of America*, 90(2):

- 874-884.
- Frigo, M., y Johnson, S. G. 2005. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2): 216–231.
- Funkhouser, T., Tsingos, N., Carlbom, I., Elko, G., Sondhi, M., West, J. E., Pingali, G., Min, P., y Ngan, A. 2004. A beam tracing method for interactive architectural acoustics. *The Journal of the Acoustical Society of America*, 115(2): 739756.
- Garas, J. 2000. *Adaptive 3D sound systems*. Springer, Netherlands.
- Gardner, W. G. 1995. Efficient convolution without input-output delay. *Journal of the Audio Engineering Society*, 43(3): 127–136.
- Gerzon, M. A. 1976. Unitary (energy-preserving) multichannel networks with feedback. *Electronics Letters*, 12(11): 278-279.
- Ginn, K. B. 1978. *Application of B&K equipment to architectural acoustics*. Brüel & Kjær.
- Giron, F. 1996. *Investigations about the directivity of sound sources*. Doctoral thesis. Ruhr-Universität Bochum, Germany.
- Goldstein, E. B. 2005. *Sensación y percepción*. Thomson, México.
- Gomes de Queiroz, M., Iazzetta, F., Kon, F., de Avelar Gomes, M., Figueiredo, F. L., Masiero, B., Ueda, L. K., Dias, L., Torres, M. H. C., y Thomaz, L. F. 2008. AcMus: an open, integrated platform for room acoustics research. *Journal of the Brazilian Computer Society*, 14(3): 87–103.
- Grantham. 1995. Spatial Hearing and Related Phenomena. En B. C. J. Moore (Ed.), *Hearing, Handbook of perception and cognition* (2nd ed.): 297-345. Academic Press, San Diego, CA.
- Haines, E. 1994. Point in Polygon Strategies. En P. Heckbert (Ed.), *Graphics Gems IV* (1st ed.): 24-46. Academic Press, Boston, MA.
- Haneda, Y., Makino, S., Kaneda, Y., y Kitawaki, N. 1999. Common-acoustical-pole and zero modeling of head-related transferfunctions. *IEEE Transactions on Speech and Audio Processing*, 7(2): 188-196.

- Hebrank, J., y Wright, D. 1974. Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America*, 56(6): 1829-1834.
- Hershkowitz, R. M., y Durlach, N. I. 1969. Interaural Time and Amplitude jnds for a 500-Hz Tone. *The Journal of the Acoustical Society of America*, 46(6B): 1464-1467.
- Hidaka, T., Yamada, Y., y Nakagawa, T. 2007. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *The Journal of the Acoustical Society of America*, 122(1): 326-332.
- Hirata, Y. 1982. A method of eliminating noise in power responses. *Journal of Sound and Vibration*, 82: 593-595.
- Hu, H., Zhou, L., Zhang, J., Ma, H., y Wu, Z. 2006. Head related transfer function personalization based on multiple regression analysis. *Proceedings of IEEE Computational Intelligence and Security*: 1829-1832. Guanzhou, China.
- Hu, H., Zhou, Lin, Ma, Hao, y Wu, Zhenyang. 2008. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, 69(2): 163-172.
- Huang, Q., y Fang, Y. 2009. Modeling personalized head-related impulse response using support vector regression. *Journal of Shanghai University (English Edition)*, 13(6): 428-432.
- Huopaniemi, J., Savioja, L., y Karjalainen, M. 1997. Modeling of reflections and air absorption in acoustical spaces a digital filter design approach: 4 pp. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
- Huopaniemi, J., y Karjalainen, M. 1996. Comparison of digital filter design methods for 3-D sound. *Proceedings IEE Nordic Signal Processing Symp. (NORSIG' 96)*: 131-134. Espoo, Finland.
- Hwang, S., Park, Y., y Park, Y. 2008. Modeling and customization of head-related impulse responses based on general basis functions in time domain. *Acta Acustica united with Acustica*, 94(6): 965-980.

- Hwang, S., y Park, Y. 2007. HRIR customization in the median plane via principal components analysis. *Proceedings of AES 31st Int. Conf. New Directions in High Resolution Audio*: 12. London, UK.
- Hwang, S., y Park, Y. 2008. Interpretations on principal components analysis of head-related impulse responses in the median plane. *The Journal of the Acoustical Society of America*, 123(4): EL65-EL71.
- Iazzetta, F., Kon, F., Gomes de Queiroz, M., Soares Correa da Silva, F., da Silva, C., y de Avelar Gomes, M. 2004. AcMus: Computational tools for measurement, analysis and simulation of room acoustics. *Proceedings of the European Acoustics Symposium*, Portugal.
- Iida, K., Itoh, M., Itagaki, A., y Morimoto, M. 2007. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, 68(8): 835-850.
- ISO. 1993. ISO 9613-1:1993 - Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere.
- ISO. 1996. ISO 10534-1:1996 - Acoustics - Determination of sound absorption coefficient and impedance in impedance tubes - Part 1: Method using standing wave ratio.
- ISO. 2003. ISO 354:2003 - Acoustics - Measurement of sound absorption in a reverberation room.
- ISO. 2009. ISO 3382-1:2009 - Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces.
- Jordan, V. L. 1981. A group of objective acoustical criteria for concert halls. *Applied Acoustics*, 14(4): 253-266.
- Jot, J. M. 1992. An analysis/synthesis approach to real-time artificial reverberation. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92*: 221-224.
- Jot, J. M., y Chaigne, A. 1991. Digital Delay Networks for Designing Artificial Reverberators. *Preprints of the Audio Engineering Society 90th Convention*:

3030. Paris, France.
- Jot, J.-M., Cerveau, L., y Warusfel, O. 1997. Analysis and synthesis of room reverberation based on a statistical time-frequency model. *Preprints of the Audio Engineering Society 103rd Convention*: 4629. NY.
- Kahana, Y., y Nelson, P. A. 2006. Numerical modelling of the spatial acoustic response of the human pinna. *Journal of Sound and Vibration*, 292(1-2): 148-178.
- Kajastila, R., Lokki, T., Lundén, P., Savioja, L., y Siltanen, S. 2007. A Distributed Real-Time Virtual Acoustic Rendering System for Dynamic Geometries. *Preprints of the Audio Engineering Society 122th Convention*. Wien, Austria.
- Kato, M., Uematsu, H., Kashino, M., y Hirahara, T. 2003. The effect of head motion on the accuracy of sound localization. *Acoustical Science and Technology*, 24(5): 315–317.
- Keet, W. V. 1968. The influence of early lateral reflections on the spatial impression. *Proceedings of the 6th International Congress on Acoustics (ICA)*: E24. Tokyo, Japan.
- Kinsler, L. E., Frey, A. R., Coppens, A. B., y Sanders, J. V. 1999. *Fundamentals of Acoustics* (4th ed.). Wiley, NY.
- Kistler, D. J., y Wightman, F. L. 1992. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3): 1637-1647.
- Kleiner, M., Dalenbäck, B.-I., y Svensson, U. P. 1993. Auralization - an overview. *Journal of the Audio Engineering Society*, 41(11): 861–875.
- Knudsen, E. I. 1981. The hearing of the barn owl. *Scientific American*, 245(6): 113–125.
- Kob, M. 2002. *Physical modeling of the singing voice*. Doctoral thesis. Aachen University, Germany.
- Krokstad, A., Strom, S., y Sørsdal, S. 1968. Calculating the acoustical room

- response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1): 118-125.
- Kuhn, G. F. 1977. Model for the interaural time differences in the azimuthal plane. *The Journal of the Acoustical Society of America*, 62(1): 157-167.
- Kuhn, G. F. 1987. Physical acoustics and measurements pertaining to directional hearing. In W. Yost y G. Gourevitch (Eds.), *Directional hearing: 3-25*. Springer-Verlag, NY.
- Kulkarni, A., Isabelle, S. K., y Colburn, H. S. 1999. Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105(5): 2821-2840.
- Kulkarni, A., y Colburn, H. Steven. 2004. Infinite-impulse-response models of the head-related transfer function. *The Journal of the Acoustical Society of America*, 115(4): 1714-1728.
- Kulowski, A. 1985. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18(6): 449-469.
- Kürer, R. 1969. Zur gewinnung von eizahlkriterien bei impulsmessungeg in der raumakustik. *Acustica*, 21: 370.
- Kuttruff, H. 1995. Sound field prediction in rooms. *Proceedings of the 15th International Congress on Acoustics (ICA)*: 545-52. Trondheim, Norway.
- Kuttruff, H. 2009. *Room Acoustics* (5th ed.). Spon Press, London.
- Langendijk, E. H. A., y Bronkhorst, A. W. 2000. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *The Journal of the Acoustical Society of America*, 107(1): 528-537.
- Lee, J. C. 2008. Hacking the Nintendo Wii Remote. *IEEE Pervasive Computing*, 7(3): 39-45.
- Lehmann, E. A., y Johansson, A. M. 2010. Diffuse Reverberation Model for Efficient Image-Source Simulation of Room Impulse Responses. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6): 1429-1439.
- Lentz, T. 2007. *Binaural technology for virtual reality*. Doctoral thesis. Aachen

- University, Germany.
- Lentz, T., Schröder, D., Vorländer, M., y Assenmacher, I. 2007. Virtual reality system with integrated sound field simulation and reproduction. *EURASIP Journal on Advances in Signal Processing*: 1-17.
- Levinson, N. 1947. The Wiener RMS error criterion in filter design and prediction. *Journal of Mathematical Physics*, 25: 261-278.
- Lindau, A. 2009. The Perception of System Latency in Dynamic Binaural Synthesis. *Proceedings of NAG/DAGA*, Rotterdam, The Netherlands.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., y Guzman, S. J. 1999. The precedence effect. *The Journal of the Acoustical Society of America*, 106(4): 1633-1654.
- Litovsky, R. Y., y Clifton, R. K. 1992. Use of sound-pressure level in auditory distance discrimination by 6-month-old infants and adults. *The Journal of the Acoustical Society of America*, 92(2 Pt 1): 794-802.
- Lokki, T., y Hiipakka, J. 2001. A time-variant reverberation algorithm for reverberation enhancement systems. *Proceedings of COST G-6 Conference on Digital Audio Effects*. Limerick, Ireland.
- Loomis, J. M., Klatzky, R. L., Philbeck, J. W., y Golledge, R. G. 1998. Assessing auditory distance perception using perceptually directed action. *Perception and Psychophysics*, 60(6): 966-980.
- Lopez-Poveda, E. A., y Meddis, R. 1996. A physical model of sound diffraction and reflections in the human concha. *The Journal of the Acoustical Society of America*, 100(5): 3248.
- Lundeby, A., Vigran, T. E., Bietz, H., y Vorländer, M. 1995. Uncertainties of Measurements in Room Acoustics. *Acta Acustica united with Acustica*, 81(4): 344-355.
- Lyon, R. H., y DeJong, R. G. 1995. *Theory and application of statistical energy analysis*. Butterworth-Heinemann, Newton, MA.
- Mackensen, P. 2004. *Auditive Localization. Head movements, an additional cue in*

- Localization*. Doctoral Thesis. Technische Universität, Berlin.
- Mankovsky, V. S. 1971. *Acoustics of studios and auditoria*. Communication Arts Books, NY.
- Mantegazza, P., Dozio, E. L., y Papacharalambous, S. 2000. RTAI: Real Time Application Interface. *Linux Journal*, 2000(72es): 10.
- Martellotta, F. 2010. The just noticeable difference of center time and clarity index in large reverberant spaces. *The Journal of the Acoustical Society of America*, 128(2): 654-663.
- Martens, W. L. 1987. Principal components analysis and resynthesis of spectral cues to perceived direction. *Proceedings of the 1987 International Computer Music Conference*: 274-281.
- Martín, R. S., Witew, I. B., Arana, M., y Vorländer, M. 2007. Influence of the Source Orientation on the Measurement of Acoustic Parameters. *Acta Acustica united with Acustica*, 93(3): 387-397.
- Meesawat, K., y Hammershoi, D. 2002. An investigation on the transition from early reflections to a reverberation tail in a BRIR. *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*. Kyoto, Japan.
- Mehrgardt, S., y Mellert, V. 1977. Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America*, 61(6): 1567-1576.
- Meyer, J. 2009. *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers* (5th ed.). Springer, NY.
- Middlebrooks, J. C. 1999. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, 106(3): 1493-1510.
- Miles, R. N. 1984. Sound field in a rectangular enclosure with diffusely reflecting boundaries. *Journal of Sound and Vibration*, 92(2): 203-226.
- Mills, A. W. 1958. On the Minimum Audible Angle. *The Journal of the Acoustical*

- Society of America*, 30(4): 237-246.
- Mills, A. W. 1960. Lateralization of High-Frequency Tones. *The Journal of the Acoustical Society of America*, 32: 132-134.
- Mills, A. W. 1972. Auditory localization. IE J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory Vol. 2*: 301-345. Academic Press, NY.
- Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., y Alloza, D. 2007. Progressive perceptual audio rendering of complex scenes. *Proceedings of the 2007 symposium on Interactive 3D graphics and games*: 189-196. ACM, NY.
- Moore, B. C. J., Oldfield, S. R., y Dooley, G. J. 1989. Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. *The Journal of the Acoustical Society of America*, 85(2): 820-836.
- Moore, B. C. J., y Glasberg, B. R. 1996. A Revision of Zwicker's Loudness Model. *Acta Acustica united with Acustica*, 82(2): 335-345.
- Moore, J. K., Perazzo, L. M., y Braun, A. 1995. Time course of axonal myelination in the human brainstem auditory pathway. *Hearing Research*, 87(1-2): 21-31.
- Moorer, J. A. 1979. About This Reverberation Business. *Computer Music Journal*, 3(2): 13-28.
- Morgan, D. R. 1997. A parametric error analysis of the backward integration method for reverberation time estimation. *The Journal of the Acoustical Society of America*, 101(5): 2686-2693.
- Morimoto, M. 2001. The contribution of two ears to the perception of vertical angle in sagittal planes. *The Journal of the Acoustical Society of America*, 109(4): 1596-1603.
- Morimoto, M., Nakagawa, K., y Iida, K. 2008. The relation between spatial impression and the law of the first wavefront. *Applied Acoustics*, 69(2): 132-140.
- Müller-Tomfelde, C. 2001. Time-Varying Filter in Non-Uniform Block

- Convolution. *Proceedings of the COST G-6 Conference on Digital Audio Effects*. Limerick, Ireland.
- Naylor, G. M. 1993. ODEON-Another hybrid room acoustical model. *Applied Acoustics*, 38(2-4): 131-143.
- Nishino, T., Kajita, S., Takeda, K., y Itakura, F. 1999. Interpolating head related transfer functions in the median plane. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*: 167-170.
- Noisternig, M., Katz, B. F. G., Siltanen, S., y Savioja, L. 2008. Framework for Real-Time Auralization in Architectural Acoustics. *Acta Acustica united with Acustica*, 94: 1000-1015.
- Novo, P. 2005. Auditory Virtual Environments. En J. Blauert (Ed.), *Communication acoustics*: 277-297. Springer, Berlin.
- Okano, T. 1998. Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. *The Journal of the Acoustical Society of America*, 104(1): 255-265.
- Okano, T. 2002. Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections. *The Journal of the Acoustical Society of America*, 111(1): 217-229.
- Oppenheim, A. V., y Schafer, R. W. 1975. *Digital signal processing*. Prentice-Hall, New Jersey.
- Pellegrini, R. S. 2002. *A virtual reference listening room as an application of auditory virtual environments*. Doctoral thesis. Ruhr-University, Bochum.
- Perrott, D. R. 1969. Rôle of Signal Onset in Sound Localization. *The Journal of the Acoustical Society of America*, 45(2): 436-445.
- Perrott, D. R. 1977. Minimum auditory movement angle: Binaural localization of moving sound sources. *The Journal of the Acoustical Society of America*, 62(6): 1463-1466.
- Perrott, D. R., Ambarsoom, H., y Tucker, J. 1987. Changes in head position as a measure of auditory localization performance: Auditory psychomotor

- coordination under monaural and binaural listening conditions. *The Journal of the Acoustical Society of America*, 82(5): 1637-1645.
- Perrott, D. R., Saberi, K., Brown, K., y Strybel, T. Z. 1990. Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, 48(3): 214-226.
- Perrott, D. R., y Saberi, K. 1990. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4): 1728-1731.
- Plenge, G. 1974. On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, 56(3): 944-951.
- Racciu, G., y Mantegazza, P. 2006. *RTAI 3.4 User Manual*. <http://www.rtai.org>
- Ramos, O. A., Araneda, M., y Tommasini, F. C. 2008. Análisis de Patrones de las Funciones de Transferencia de Cabeza. *Actas del VI Congreso Iberoamericano de Acústica - FIA 2008*, Buenos Aires.
- Ramos, O. A., Araneda, M., y Tommasini, F. C. 2009a. Diseño y evaluación objetiva de filtros binaurales. *Mecánica Computacional*, XXVIII: 137-148.
- Ramos, O. A., Araneda, M., y Tommasini, F. C. 2009b. Construcción de HRTF individuales a partir de mediciones antropométricas. *Mecánica Computacional*, XXVIII: 149-161.
- Ramos, O. A., Arias, C., Tommasini, F. C., Ortiz Skarp, A. H., Novillo, D., Bermejo, F., Hüg, M. X, Lunati, V. 2009. Movimientos de cabeza en una prueba de ecolocación. *Mecánica Computacional*, XXVIII: 123-135.
- Ramos, O. A., Calvo, G., y Tommasini, F. C. 2007. Modelo Acústico de Cabeza y Torso Mediante Análisis de Componentes Principales. *Mecánica Computacional*, XXVI: 46-58.
- Ramos, O. A., Tommasini, F. C., y Araneda, M. 2010. Contribución de la cabeza, el torso y el oído externo en las funciones de transferencia relativas a la cabeza. *Actas del 2º Congreso Internacional de Acústica UNTREF*. Buenos Aires, Argentina.

- Ramos, O. A., y Tommasini, F. C. 2008. Simplificación de las funciones de transferencia de cabeza mediante análisis de componentes principales. *Mecánica Computacional*, XXVII: 431–442.
- Reichardt, W., Alim, O. A., y Schmidt, W. 1975. Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances. *Acustica*, 32: 126–137.
- Rife, D. D., y Vanderkooy, J. 1989. Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*, 37(6): 419–444.
- Rindel, J. H. 2000. The use of computer modeling in room acoustics. *Journal of Vibroengineering*, 3(4): 41–72.
- Rindel, J. H., Otondo, F., y Christensen, C. L. 2004. Sound source representation for auralization. *Proceedings of International Symposium on Room Acoustics: Design and Science*, Hyogo, Japan.
- Riva, M. H. 2009. Síntesis en Tiempo Real de Sonido Tridimensional sobre Auriculares. *RPIC Estudiantil 2009*: 1100-1106. Rosario.
- Rodríguez, S. G., y Ramírez, M. A. 2005. Linear relationships between spectral characteristics and anthropometry of the external ear. *Proceedings of 11th Meeting of the International Conference on Auditory Display*: 336-339, Limerick, Ireland.
- Rosenblum, L. D., Wuestefeld, A. P., y Anderson, K. L. 1996. Auditory Reachability: An Affordance Approach to the Perception of Sound Source Distance. *Ecological Psychology*, 8(1): 1-24.
- Rubak, P., y Johansen, L. G. 1999. Artificial Reverberation Based on a Pseudo-Random Impulse Response II. *Preprints of the Audio Engineering Society 106th Convention*, Munchen, Germany.
- Sanderson, C. 2010. *Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments*. Technical Report. Australia: NICTA.
- Sandvad, J. 1996. Dynamic Aspects of Auditory Virtual Environments. *Preprints*

- of the Audio Engineering Society 100th Convention*, Copenhagen, Denmark.
- Sandvad, J., y Hammershoi, D. 1994. Binaural Auralization, Comparison of FIR and IIR Filter Representation of HIRs. *Preprints of the Audio Engineering Society 96th Convention*, Amsterdam.
- Savioja, L. 1999. *Modeling techniques for virtual acoustics*. Doctoral thesis. Helsinki University of Technology, Finland.
- Savioja, L., Huopaniemi, J., Lokki, T., y Väänänen, R. 1999. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9): 675–705.
- Savioja, L., Manocha, D., y Lin, M. 2010. Use of GPUs in room acoustic modeling and auralization. *Proceedings of International Symposium on Room Acoustics. ISRA*. Melbourne.
- Scarpaci, J. W. 2006. *Creation of a system for real time virtual auditory space and its application to dynamic sound localization*. Doctoral thesis. Boston University, Boston, MA.
- Scarpaci, J. W., y Colburn, H. S. 2005. Principal components analysis interpolation of head related transfer functions using locally-chosen basis functions. *Proceedings of 11 Meeting of the International Conference on Auditory Display*. Limerick, Ireland.
- Scavone, G. P., y Cook, P. R. 2005. RtMidi, RtAudio, and a Synthesis Toolkit (STK) update. *Proceedings of the 2005 International Computer Music Conference*: 1-4. Barcelona, España.
- Schroeder, M. R. 1961. Novel Uses of Digital Computers in Room Acoustics. *The Journal of the Acoustical Society of America*, 33(11): 1669.
- Schroeder, M. R. 1962a. Natural sounding artificial reverberation. *Journal of the Audio Engineering Society*, 10(3): 219–223.
- Schroeder, M. R. 1962b. Frequency-Correlation Functions of Frequency Responses in Rooms. *The Journal of the Acoustical Society of America*, 34: 1819-1823.
- Schroeder, M. R. 1965. New method of measuring reverberation time. *The Journal*

- of the Acoustical Society of America*, 37(3): 409–412.
- Schroeder, M. R. 1973. Computer Models for Concert Hall Acoustics. *American Journal of Physics*, 41(4): 461.
- Schroeder, M. R., Gottlob, D., y Siebrasse, K. F. 1974. Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. *The Journal of the Acoustical Society of America*, 56(4): 1195-1201.
- Shaw, E. A. . 1997. Acoustical Features of the Human External Ear. En R. H. Gilkey y T. R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum, Mahwah, NJ.
- Shaw, E. A. G. 1974. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America*, 56(6): 1848-1861.
- Shaw, E. A. G., y Teranishi, R. 1968. Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source. *The Journal of the Acoustical Society of America*, 44(1): 240-249.
- Shlens, J. 2005. *A tutorial on principal component analysis*. Technical Report, Systems Neurobiology Laboratory, University of California, San Diego.
- Silzle, A., Strauss, H., y Novo, P. 2004. IKA-SIM: A System to Generate Auditory Virtual Environments. *Preprints of the Audio Engineering Society 116th Convention*: 6016.
- Simpson, W. E., y Stanton, L. D. 1973. Head Movement Does Not Facilitate Perception of the Distance of a Source of Sound. *The American Journal of Psychology*, 86(1): 151-159.
- Smith, J. O. 2010. *Physical Audio Signal Processing: for Virtual Musical Instruments and Digital Audio Effects*. W3K Publishing, Stanford.
- Soulodre, G. A., y Bradley, J. S. 1995. Subjective evaluation of new room acoustic measures. *The Journal of the Acoustical Society of America*, 98(1): 294-301.
- Speigle, J. M., y Loomis, J. M. 1993. Auditory distance perception by translating

- observers: 92–99. IEEE, San Jose, CA.
- Stautner, J., y Puckette, M. 1982. Designing Multi-Channel Reverberators. *Computer Music Journal*, 6(1): 52-65.
- Steiglitz, K., y McBride, L. 1965. A technique for the identification of linear systems. *IEEE Transactions on Automatic Control*, 10(4): 461–464.
- Stewart, R., y Sandler, M. 2007. Statistical measures of early reflections of room impulse responses. *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*: 59–62. Bordeaux, France.
- Tang, S. K., y Yeung, M. H. 2006. Reverberation times and speech transmission indices in classrooms. *Journal of Sound and Vibration*, 294(3): 596-607.
- Teranishi, R., y Shaw, E. A. G. 1968. External-Ear Acoustic Models with Simple Geometry. *The Journal of the Acoustical Society of America*, 44(1): 257-263.
- Thiele, R. 1953. Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen [Direction and time dependence of sound reflections in rooms]. *Acustica*, 3: 291-302.
- Thurlow, W. R., Mangels, J. W., y Runge, P. S. 1967. Head Movements During Sound Localization. *The Journal of the Acoustical Society of America*, 42(2): 489-493.
- Tollin, D. J. 2004. Development of the sound localization cues in cats. *The Journal of the Acoustical Society of America*, 115(5): 2535.
- Tommasini, F. C., Ramos, O. A., y Ferreyra, S. 2008a. Modelado eficiente y validación objetiva de la respuesta impulsiva de un recinto. *Mecánica Computacional*, XXVII: 443-455.
- Tommasini, F. C., Ramos, O. A., y Ferreyra, S. 2008b. Comparación objetiva de respuestas impulsivas de recintos medidas y simuladas. *VI Congreso Iberoamericano de Acústica (FIA 2008)*, Buenos Aires, Argentina.
- Tommasini, F. C., y Ramos, O. A. 2008. Modelo basado en evidencias perceptuales para simular respuestas impulsivas de recintos. *Actas de la XXXIV*

- Conferencia Latinoamericana de Informática (CLEI 2008)*: 1100-1109. Santa Fe, Argentina.
- Tommasini, F. C., y Ramos, O. A. 2010. Auralización en tiempo real: implementación del modelo del oyente. *Actas del 2º Congreso Internacional de Acústica UNTREF*. Buenos Aires, Argentina.
- Torger, A., y Farina, A. 2001. Real-time partitioned convolution for Ambiophonics surround sound. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*: 195-198, New Paltz, NY.
- Torres, R. R., de Rycker, N., y Kleiner, M. 2004. Edge diffraction and surface scattering in concert halls: physical and perceptual aspects. *Journal of Temporal Design in Architecture and the Environment*, 4(1): 52-58.
- Torres, R. R., Svensson, U. P., y Kleiner, M. 2001. Computation of edge diffraction for more accurate room acoustics auralization. *The Journal of the Acoustical Society of America*, 109(2): 600-610.
- Tsingos, N., Jiang, W., y Williams, I. 2011. Using Programmable Graphics Hardware for Acoustics and Audio Rendering. *Journal of the Audio Engineering Society*, 59(9): 628-646.
- Tsingos, N., y Gascuel, J. D. 1997. *Acoustic simulation using hierarchical time-varying radiant exchanges*. Unpublished technical Report. iMAGIS-GRAVIR/IMAG.
- Van Maercke, D. 1986. Simulation of sound fields in time and frequency domain using a geometrical model. *Proceedings of the 12th International Congress of Acoustics (ICA '86)*. Toronto, Canada.
- Vanderkooy, J. 1994. Aspects of MLS measuring systems. *Journal of the Audio Engineering Society*, 42(4): 219-231.
- Vorländer, M. 2007. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (1st ed.). Springer, Berlin.
- Vorländer, M., y Bietz, H. 1994. Comparison of methods for measuring reverberation time. *Acustica*, 80(3): 205-215.

- Vun, N., Hor, H. F., y Chao, J. W. 2008. Real-Time Enhancements for Embedded Linux. *14th IEEE International Conference on Parallel and Distributed Systems*: 737–740.
- Wallach, H. 1939. On Sound Localization. *The Journal of the Acoustical Society of America*, 10(4): 270-274.
- Wallach, H. 1940. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4): 339–368.
- Wallach, H., Newman, E. B., y Rosenzweig, M. R. 1949. The precedence effect in sound localization. *The American Journal of Psychology*, 62(3): 315–336.
- Wefers, F., y Schröder, D. 2009. Real-time auralization of coupled rooms. *Proceedings of the EAA Symposium on Auralization*. Espoo, Finland.
- Wenzel, E. M. 1998. The impact of system latency on dynamic performance in virtual acoustic environments. *Proceedings of the 16th International Congress of Acoustics and 135th Meeting of the Acoustical Society of America*: 180. Seattle, WA.
- Wenzel, E. M. 1999. Effect of Increasing System Latency on Localization of Virtual Sounds. *Proceedings of the Audio Engineering Society 16th International Conference on Spatial Sound Reproduction: 42-50*, New York.
- Wenzel, E. M. 2001. Effect of increasing system latency on localization of virtual sounds with short and long duration. *Proceedings of International conference on Auditory Display (ICAD)*. Espoo, Finland.
- Wenzel, E. M., Arruda, M., Kistler, D. J., y Wightman, F. L. 1993. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1): 111-123.
- Wenzel, E. M., Miller, J. D., y Abel, J. S. 2000. Sound Lab: A Real-Time, Software-Based System for the Study of Spatial Hearing. *Preprints of the Audio Engineering Society 108th Convention*. Paris, France.
- Wenzel, E. M., y Foster, S. H. 1993. Perceptual consequences of interpolating head-related transferfunctions during spatial synthesis. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*: 102-105.

- Wightman, F. L., Kistler, D. J., y Arruda, M. 1992. Perceptual consequences of engineering compromises in synthesis of virtual auditory objects. *The Journal of the Acoustical Society of America*, 92(4): 2332.
- Wightman, F. L., y Kistler, D. J. 1989a. Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2): 858-867.
- Wightman, F. L., y Kistler, D. J. 1989b. Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, 85(2): 868-878.
- Wronski, M. P. 2008. *Design and Implementation of a Hand Tracking Interface using the Nintendo Wii Remote*. Bachelor thesis, University of Cape Town, Cape Town.
- Xiang, N. B. 1991. Computer-Aided Tenth-Scale Modeling for Binaural Auralization in Room Acoustic Design. *Peprints of the Audio Engineering Society 91st Convention*.
- Xu, S., Li, Z., y Salvendy, G. 2007. Individualization of head-related transfer function for three-dimensional virtual auditory display: a review. *Proceedings of the 2nd International Conference on Virtual Reality*: 397–407. Berlin, Heidelberg.
- Xu, S., Li, Z., y Salvendy, G. 2009. Identification of Anthropometric Measurements for Individualization of Head-Related Transfer Functions. *Acta Acustica united with Acustica*, 95(1): 168–177.
- Yaghmour, K. 2001. *Adaptive domain environment for operating systems*. Opersys.
- Yairi, S., Iwaya, Y., y Suzuki, Y. 2006. Investigation of system latency detection threshold of virtual auditory display. *Proceedings of International conference on Auditory Display (ICAD)*: 217–222.
- Yost, W. A. 1991. Auditory image perception and analysis: The basis for hearing. *Hearing Research*, 56(1-2): 8-18.
- Yost, W. A., y Soderquist, D. R. 1984. The precedence effect: Revisited. *The Journal of the Acoustical Society of America*, 76(5): 1377-1383.

- Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., y Tam, C. 2006. Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *The Journal of the Acoustical Society of America*, 120(1): 343-359.
- Zhou, B., Green, D. M., y Middlebrooks, J. C. 1992. Characterization of external ear impulse responses using Golay codes. *The Journal of the Acoustical Society of America*, 92(2): 1169-1171.