

UN MODELO PARA LA SIMULACIÓN DEL ESPACIO EN MÚSICA



Instituto de Investigación
Musicológica "Carlos Vega"



fuentes

INSTITUTO DE INVESTIGACIÓN MUSICOLÓGICA
"CARLOS VEGA"

**UN MODELO
PARA LA SIMULACIÓN
DEL ESPACIO EN MÚSICA**

PABLO CETTA



Editorial de la Universidad Católica Argentina

Cetta, Pablo

Un modelo para la simulación del espacio en música – 1a ed. – Buenos Aires : Educa, 2007.

152 p. ; 28x20 cm.

ISBN 978-987-1190-95-9

1. Música. 2. Simulación. I. Título
CDD 780.1



**EDITORIAL
DE LA UNIVERSIDAD
CATÓLICA ARGENTINA**

FUNDACIÓN UNIVERSIDAD CATÓLICA ARGENTINA
A. M. de Justo 1400 • P.B., Contrafrente • (C1107AAZ)
Tel./Fax 4349-0200 • educa@uca.edu.ar
Buenos Aires, julio de 2010

ISBN: 987-1190-95-9

Índice

Introducción		1
Capítulo 1	Factores psicofísicos que determinan la localización de los eventos sonoros	
1.0	Introducción	12
1.1	Funciones de transferencia del oído externo	15
1.2	Indicios de lateralización y elevación de la fuente	23
1.2.0	Diferencia interaural de nivel de presión sonora (ILD)	23
1.2.1	Diferencia interaural de tiempo (ITD)	29
1.2.2	Indicios espectrales monoaurales	35
1.2.3	Correcciones asociadas a los movimientos de la cabeza	46
1.3	Divergencias en la posición del evento sonoro y del evento auditivo	48
1.3.0	Localización indefinida	48
1.3.1	Bandas direccionales	50
1.3.2	Influencia del grado de coherencia de las señales	55
1.4	Percepción de la distancia	57
1.5	Audición en recintos cerrados	59
1.6	Conclusiones	64
Capítulo 2	Simulación de fuentes aparentes en ambientes virtuales	
2.0	Introducción	67
2.1	Simulación a partir de la estereofonía	68
2.1.0	Fundamentos del sistema estereofónico	68
2.1.1	Un modelo simple de localización espacial	70
2.2	Extensión al sistema cuadrafónico	77
2.2.0	Descripción del modelo de Chowning	77
2.2.1	Implementación en tiempo real	79
2.3	Ambientes virtuales	81
2.3.0	Simulación de las primeras reflexiones	82
2.3.1	Simulación de la reverberación	85
2.3.2	Descripción del modelo de Moore	96
2.3.3	Implementación en tiempo real	99
2.4	Control vectorial de la amplitud sobre ternas de parlantes	103
2.4.0	Descripción del modelo de Pulkki	103
2.4.1	Implementación en tiempo real	107
2.5	Sistemas basados en la reconstrucción del frente de onda	109
2.5.0	Wave Field Synthesis	109
2.5.1	Ambisonics	111
2.5.2	Implementación en tiempo real	122
2.6	Conclusiones	126

Capítulo 3	Diseño de programas asistentes para el tratamiento del espacio y composición musical	
3.0	Introducción	130
3.1	Técnicas binaurales	131
3.1.0	Principios de grabación binaural	131
3.1.1	Síntesis binaural	132
3.1.2	HRTF medidas	132
3.1.3	HRTF modeladas	134
3.1.4	Aplicación de la convolución con HRTF medidas	136
3.2	Implementación en un primer programa	140
3.2.0	Acerca de las HRTF utilizadas	140
3.2.1	Preparación de las HRTF	142
3.2.2	Principios de funcionamiento del software	142
3.2.3	Versión para OS Windows	145
3.3	Un sistema de localización espacial en tiempo real	146
3.3.0	Síntesis binaural con Max-MSP	146
3.3.1	Simulación de la absorción del aire	147
3.3.2	Simulación de las primeras reflexiones	147
3.3.3	Descripción de la posición de la fuente	150
3.3.4	Aplicación de transformaciones geométricas	151
3.3.5	Desarrollo de una interfaz gráfica	157
3.3.6	Reproducción transaural	160
3.3.7	Extensión a la cuadrafonía	166
3.3.8	Configuraciones posibles	167
3.4	Integración de la música al espacio virtual	168
3.4.0	Exteriorización e interiorización del espacio	169
3.4.1	Tratamiento espacial de la altura y el timbre	172
3.4.2	Tratamiento espacial de las relaciones temporales	175
3.4.3	Tratamiento espacial de la textura	179
3.5	Descripción general de la obra	185
	Partitura de <i>Interiores</i>	188
3.6	Conclusiones	202
	Referencias bibliográficas	204

Lista de figuras

1.1	Planos relativos al oyente y ángulos de posicionamiento de la fuente	14
1.2	Función de transferencia en un sistema lineal	15
1.3	Filtro caracterizado por su respuesta a impulso	16
1.4	Respuestas a impulso registradas con una cabeza artificial para $\theta = 30^\circ$ y $\varphi = 0^\circ$.	17
1.5	HRTF medidas sobre dos sujetos entre 0° y 30° de elevación, en pasos de 10° (Kendall y Martens, 1984)	19
1.6	Funciones de transferencia <i>free-field</i> para $\theta = 0^\circ$ y $\varphi = 0^\circ$ (Blauert, 1997)	21
1.7	Funciones de transferencia monoaurales para una referencia de $\theta = 0^\circ$ y $\varphi = 0^\circ$ (Blauert, 1997)	22
1.8	Diferencia calculada de nivel de presión sonora entre un punto de una esfera rígida, ubicado a 90° , y <i>free-field</i> . (Shaw, 1974)	24
1.9	Diferencia interaural de nivel de presión sonora, calculada sobre un modelo de esfera rígida. (Blauert, 1997)	25
1.10	Diferencias interaurales de nivel de presión sonora para distintas bandas de frecuencia, medidas sobre un sujeto. (Wightman y Kistler, 1997)	26
1.11	Transformaciones <i>free-field</i> promedio, sobre el plano horizontal, a intervalos de 15° , de una síntesis de doce estudios. (Shaw, 1974)	27
1.12	Desplazamiento lateral del evento auditivo, medido en función del retardo interaural. (Toole y Sayers, 1965)	30
1.13	Curva de lateralización por retardos para una señal sinusoidal de 600 Hz. (Sayers, 1964)	31
1.14	Curva de lateralización por retardos para señales distintas. (Scherer, 1959)	31
1.15	Relación entre ITD e ILD para distintas sonoridades. (David, Guttman y van Bergeijk, 1959)	34
1.16	Cono de confusión	34
1.17	Funciones del pabellón auditivo, según Batteau (1967)	36
1.18	Resonancias del pabellón auditivo (Shaw y Tranishi, 1968).	37
1.19	Errores en la localización generados por alteraciones sistemáticas del espectro de la fuente (Wightman y Kistler, 1997).	39
1.20	DTFs de un sujeto, medidas a 90° de azimut (Wightman y Kistler, 1997).	40
1.21	DTFs de un sujeto, medidas sobre el plano horizontal y el plano medio (Middlebrooks, 1997).	41
1.22	Conos de azimut constante (Duda, 1997).	42
1.23	Diferencias de amplitud y tiempo medidas a $\theta = 30^\circ$ y $\varphi = 0^\circ$, con cabeza artificial (Duda, 1997).	43
1.24	Diferencia interaural de intensidad a 40° de azimut y ángulos entre -90° y 230° de elevación, sobre una cabeza artificial (Duda, 1997).	44
1.25	Curva suavizada de diferencia interaural de intensidad, para apreciar el contorno casi sinusoidal (Duda, 1997).	45
1.26	Indicio aportado por el movimiento de la cabeza.	46

1.27	Percepción simétrica respecto al eje interaural	47
1.28	Ángulo mínimo audible en función de la frecuencia (Mills, 1958).	49
1.29	Localización del evento auditivo en función de su frecuencia (Roffler y Butler, 1968)	51
1.30	Bandas direccionales (Blauert, 1997)	52
1.31	Nivel de presión sonora medido en el tímpano con señales al frente (<i>v</i>) y arriba (<i>o</i>), respecto a la presión medida atrás (Blauert, 1997)	53
1.32	Picos encubiertos. Magnitud de la HRTF vs. azimut y vs. frecuencia, con frecuencia y posición como parámetros, respectivamente (Butler, 1987)	54
1.33	Influencia del nivel de coherencia en la ubicación de los eventos auditivos (Chernyak y Dubrovsky, 1968)	55
1.34	Influencia de la posición de dos fuentes incoherentes en la ubicación de los eventos auditivos (Damaske, 1967/68)	56
1.35	Evaluación de la distancia subjetiva en función de la distancia del evento sonoro, para la voz hablada (Gardner, 1969)	57
1.36	Relación entre la distancia estimada y el nivel de la señal, para dos posiciones fijas de 3 y 9 metros (Gardner, 1969).	58
1.37	Nivel de atenuación del aire, en función de la frecuencia (Blauert, 1997).	58
1.38	Respuesta típica de una sala.	60
2.1	Sistema estereofónico. Imagen sonora generada por diferencias de intensidad y retardos.	68
2.2	Sistema estereofónico. Técnicas microfónicas M-S y ORTF.	69
2.3	Espacio virtual generado en la simulación.	70
2.4	Formalización de las diferencias de distancia entre la fuente y cada oído.	73
2.5	Interfaz gráfica del modelo simple implementado en Csound (Cetta, 1996).	76
2.6	Esquema del sistema de localización de J. Chowning.	77
2.7	Función alojada en un oscilador que evita el cálculo de las ganancias.	78
2.8	El objeto <i>Turenas</i> , que contiene al espacializador (Cetta, 2003a).	80
2.9	El objeto <i>Turenas</i> , por dentro (Cetta, 2003a).	80
2.10	Unidad de lectura de los coeficientes de amplitud del objeto <i>Turenas</i> (Cetta, 2003a).	81
2.11	Método geométrico de las fuentes imaginarias para el cálculo de las reflexiones en un recinto.	83
2.12	Línea de retardo.	83
2.13	Simulación de las primeras reflexiones con líneas de retardo en serie.	84
2.14	Filtro <i>feedforward comb</i> .	85
2.15	Filtro <i>feedback comb</i> .	86
2.16	Filtro <i>allpass</i> .	87
2.17	Ejemplo de filtros <i>allpass</i> en cascada.	88
2.18	Filtro <i>comb</i> con pasabajos.	89
2.19	Reverberador propuesto por Moorer (1979), con 6 <i>comb</i> y 1 <i>allpass</i> .	90
2.20	Reverberador <i>JCRev</i> , de Chowning.	91

2.21	Estructura FDN para reverberación propuesta por Jot (1991).	92
2.22	Ambiente real y ambiente virtual en el modelo de Moore (1983).	97
2.23	Primeras reflexiones que alcanzan la posición de los parlantes frontales.	98
2.24	Aplicación del método de las fuentes imaginarias en el cálculo de las distancias fuente-parlante, para las señales reflejadas.	98
2.25	Sección de cálculo de los tiempos de retardo y amplitudes en la versión en tiempo real (Cetta, 2003a).	100
2.26	Aplicación práctica de los retardos y ganancias en la versión en tiempo real del modelo de Moore (Cetta, 2003a).	101
2.27	Subdivisión de un sector de la esfera virtual que rodea al oyente en ternas de parlantes.	103
2.28	El sistema estereofónico definido vectorialmente.	104
2.29	Triángulo activo definido por tres canales.	105
2.30	Ángulo de expansión en el método MDAP (Pulkki, 1999).	106
2.31	Inicialización del objeto <i>define_loudspeakers</i> (Pulkki, 2000).	107
2.32	Implementación del método MDAP en 3D (Pulkki, 2000).	108
2.33	Unidad de retardos para compensar las diferencias entre las distancias a los parlantes (Pulkki, 2000).	108
2.34	Reconstrucción del frente de onda por medio de <i>wavelets</i> .	109
2.35	Detalle interior del micrófono <i>Calrec Soundfield</i> .	114
2.36	Esquema de implementación del sistema Ambisonics.	117
2.37	Representación de los armónicos esféricos.	121
2.38	Interfaz gráfica par el trazado de trayectorias en el programa WDSPE (Di Liscia, 1999).	121
2.39	Codificador Ambisonics de primer orden en Max-MSP.	122
2.40	Decodificador Ambisonics de primer orden, para cuatro parlantes dispuestos horizontalmente.	122
2.41	Decodificador Ambisonics de primer orden, para ocho parlantes dispuestos en un cubo.	123
2.42	Codificador Ambisonics de segundo orden.	123
2.43	Decodificador Ambisonics de segundo orden para cuatro parlantes dispuestos horizontalmente.	124
2.44	Decodificador Ambisonics de segundo orden para ocho parlantes dispuestos en un cubo.	124
3.1	Grabación binaural utilizando una cabeza artificial y ecualizaciones asociadas.	131
3.2	Componentes que integran el modelado físico de las HRTF.	135
3.3	Implementación de filtro FIR para convolución lineal.	137
3.4	Implementación de filtro FIR para convolución lineal.	144
3.5	Software de localización espacial basado en síntesis binaural. Versión para sistema operativo Windows de 32 bits.	145
3.6	Archivo de ayuda del objeto de síntesis binaural, programado para el entorno Max-MSP.	146
3.7	Objeto de simulación de la absorción del aire, programado para el entorno Max-MSP.	147
3.8	Pantalla de ayuda del objeto que calcula la posición de las fuentes fantasma, destinadas a la simulación de las primeras reflexiones.	148

3.9	Pantalla de ayuda del objeto que calcula la distancia de una fuente fantasma, y aplica transformaciones a la señal de audio, en relación a esa distancia.	148
3.10	Pantalla de ayuda del objeto que procesa el sonido directo y las primeras reflexiones, dadas la posición de la fuente y las dimensiones del ambiente.	149
3.11	Posicionamiento de la fuente a través de dos vectores solidarios.	150
3.12	Objetos que convierten coordenadas cartesianas a esféricas, y viceversa.	151
3.13	Coordenadas homogéneas en 2D, para $M = 1$.	154
3.14	Objeto de control <i>coord</i> , que calcula la posición final de la fuente al dar datos sobre los vectores que definen su posición.	158
3.15	Pantalla de ayuda del objeto de visualización de los movimientos de la fuente <i>tr_graph</i> .	158
3.16	Pantalla principal de la interfaz gráfica para el trazado de trayectorias.	159
3.17	Objeto destinado al muestreo y almacenamiento de trayectorias espaciales.	160
3.18	Matriz para eliminación del <i>crosstalk</i> propuesta por Schroeder (1963).	161
3.19	Matriz <i>lattice</i> (a) y matriz <i>shuffler</i> (b) comparadas.	164
3.20	Implementación de matriz <i>shuffler</i> para reproducción transaural.	164
3.21	Pantalla de ayuda del objeto <i>ztransaural</i> , para reproducción transaural.	165
3.22	Deformación de la imagen transaural en un ambiente LEDE, según la posición del oyente Kendall y Martens, 1984).	166
3.23	Extensión de la reproducción transaural a la cuadrafonía, a través del objeto <i>zpan_trans~</i> .	167
3-24	Objeto que detecta cambios de altura, y los reporta, sólo si ocurren por encima de un cierto umbral de amplitud.	169
3.25	Esquema de implementación de módulo de espacialización en función de la altura, la dinámica y el grado de tonicidad de los eventos sonoros.	171
3.26	Relación entre los parámetros de un gesto musical y su desplazamiento en el espacio (a) y utilización de la misma trayectoria en el procesamiento de un objeto sonoro con un filtro (b).	171
3.27	Objeto que calcula índices de transposición de acuerdo a un PCS especificado.	174
3.28	Relaciones entre los tiempos de llegada de las primeras reflexiones, medidos con segmentos.	175
3.29	Relaciones entre los tiempos de llegada de las primeras reflexiones, variando el ángulo de azimut.	176
3.30	Ídem figura 3.29, pero representado en notación musical.	177
3.31	Software que genera ritmos a partir de trayectorias.	178
3.32	Sucesiones de sonidos breves, agrupadas de acuerdo a cantidad de alturas ejecutadas.	179
3.33	Objeto que fragmenta un evento continuante.	180
3.34	Objeto que devuelve coordenadas aleatorias dentro de un rango especificado.	180
3.35	Módulos de fragmentación (a) y transposición según PCS (b).	181
3.36	Ubicación de fuentes sonoras en complejos geométricos.	181
3.37	Pantalla principal del programa de procesamiento de <i>Interiores</i>	183

Introducción

Introducción

Esta tesis describe el diseño y desarrollo de programas de asistencia en la composición musical, especialmente orientados al tratamiento del espacio en música. El software se aplica a la creación de una obra original, centrada en la misma problemática.

La composición musical y el desarrollo de programas presentan temáticas bien diferenciadas, pero mutuamente relacionadas. El primer objetivo, la composición, surge del deseo de expresar la poética del espacio en términos sonoros. El segundo, las herramientas de creación, de la necesidad del compositor de generar un entorno apropiado que posibilite la realización de su obra, sin el cual ésta no podría existir.

Los programas asistentes, empleados en la génesis de la obra sirven a:

- la simulación de fuentes aparentes en ambientes virtuales, y su desplazamiento en el espacio, a través de métodos eficientes
- el tratamiento del espacio interno de la música en relación a las trayectorias descriptas por esas fuentes.
- el control espacial de la altura, las relaciones temporales entre los eventos sonoros, el timbre y la textura.
- la generación de volúmenes espaciales, y el control de la posición de las fuentes que los generan.

Se trata de un trabajo interdisciplinario, donde intervienen temas relacionados con la audición, el procesamiento de señales digitales, la informática y la música.

Los dos campos de estudio principales que forman la base científica de la audición espacial son la psicoacústica y la neurofisiología. El primero se concentra en las transformaciones que

afectan a las ondas que parten de una o más fuentes sonoras y llegan a los oídos, y su incidencia en las imágenes auditivas, producidas a partir de los estímulos físicos recibidos. El segundo, analiza los procesos neurológicos que intervienen en la formación de esas imágenes. Los datos que brinda la psicoacústica resultan suficientes en relación al primer objetivo planteado, por lo cual limitaremos el área de estudio relativa a la audición al campo psicofísico. Resulta pertinente, entonces, comenzar con una evaluación de sus aportes. El primer capítulo está centrado, principalmente, en el análisis de los indicios que sirven a la localización. Contempla, además, las divergencias que ocurren entre las sensaciones derivadas de la audición espacial y los estímulos que las producen, y también los fenómenos auditivos relacionados con la percepción en el interior de recintos acústicos. Toda esta información resulta necesaria si deseamos lograr una simulación efectiva de eventos auditivos, utilizando medios electroacústicos.

A fin de lograr el desarrollo de un sistema de simulación espacial aplicado a la música, conviene realizar un relevamiento previo de los sistemas existentes. La tarea no debe limitarse sólo a la descripción de las características más salientes de estos sistemas, sino al traslado de algunos de ellos al ámbito del procesamiento en tiempo real, incorporándolos como asistentes adicionales del entorno compositivo aquí propuesto. El segundo capítulo, entonces, se refiere a técnicas de simulación de imágenes espaciales en ambientes ilusorios -que emulan los indicios de localización expuestos en el capítulo anterior- y su reescritura en el dominio del procesamiento sonoro en vivo. Al campo de conocimientos psicoacústicos antes mencionado, se incorpora el procesamiento digital de señales de audio, y la programación, como medio indispensable para alcanzar los objetivos buscados. Dado que los desarrollos persiguen una aplicación en el ámbito de la composición musical, es esta disciplina la que orienta las decisiones y caminos a seguir.

El tercer y último capítulo se inicia con los lineamientos teóricos en los que se sustenta el modelo de localización espacial propuesto para esta tesis, y continúa con detalles de su implementación. Se trata de un dispositivo cuadrafónico de síntesis binaural, con

reproducción doble transaural, desarrollado en el entorno gráfico de programación de audio digital en tiempo real Max-MSP. El software creado procesa el sonido proveniente de instrumentos musicales tradicionales, o de fuentes electrónicas, y lo distribuye espacialmente a través de parlantes en la sala de concierto. Transforma las señales monofónicas que ingresan a la computadora, e induce al oyente a percibir una imagen espacial, similar a la que experimentaría frente a una o más fuentes estáticas o en movimiento, y un espacio de características materiales distintas a las del recinto en que se encuentra. El capítulo continúa con la descripción de otros programas propios, destinados a la producción de la obra, que según mencionáramos antes se destinan al control de los parámetros musicales en un contexto espacial.

La obra *Interiores*, escrita para flauta, oboe, saxo alto, clarinete bajo y piano, con procesamiento electrónico del sonido en tiempo real, explora la integración de la música al espacio virtual. Si bien debe ser considerada por sus valores poéticos también representa, en este contexto, un modo de validación de los resultados.

No estimo demasiado necesaria una justificación de las causas que guían a un compositor a desarrollar un entorno propicio para la creación, ni las razones que imponen la incursión en campos de conocimiento aparentemente alejados del quehacer musical. Una simple observación de la historia de la música puede brindar las respuestas apropiadas. No obstante, la aclaración de algunos puntos puede resultar conveniente.

A fines de la década del 50 comienzan a utilizarse las computadoras en música. Varios compositores, animados por un control exhaustivo del sonido y de las estructuras musicales, vieron en las máquinas el medio más preciso para concretar sus ideas musicales. La computadora, en su relación con la música, es un instrumento capaz de crear nuevos instrumentos, que suenan, pero que a la vez ejecutan, asisten en la composición, analizan música y sonido, perciben las intenciones de un intérprete, recuerdan, modifican el sonido de los instrumentos tradicionales, aprenden y enseñan.

De medios nuevos surgen estéticas nuevas, y muchas de ellas son motivo de confrontaciones y controversias. Lo cierto es que la música que hace uso de los medios tecnológicos envuelve una actividad multidisciplinaria, que obliga a moverse más allá de las especialidades. Los científicos actúan en el campo del arte y los artistas en el de la ciencia. Entre tanto, la música gana poco a poco un nuevo espacio.

La música electroacústica se inicia en este terreno. La primera obra que hizo uso de la espacialización del sonido es *Gesang der Jünglinge* (1955-56), de Karlheinz Stockhausen, concebida originalmente para cinco grupos de parlantes, distribuidos en las esquinas del auditorio, y en el techo. La siguiente obra electrónica del autor es *Kontakte* (1959-60), de la cual produjo luego una segunda versión, que incorpora piano y percusión combinados con los sonidos electrónicos. Para esta obra, Stockhausen diseñó un mecanismo basado en una mesa redonda giratoria, con un parlante fijado a su superficie. A los lados de la mesa ubicó cuatro micrófonos que registraban los sonidos producidos por ese parlante. Con ayuda de un cronómetro, podía girar la mesa y realizar las trayectorias circulares planeadas en su partitura. El sistema puede hoy parecerse precario, pero la audición de la obra aún nos sigue sorprendiendo.

Otro ejemplo de aplicación de la tecnología, pero en la música instrumental, lo encontramos en las obras del movimiento espectral francés, que comienza a mediados de la década del 70, y cuyos principios aún subsisten en la música de los compositores actuales. Esta música parte de la resíntesis instrumental de espectros sonoros, casi siempre tomados del sonido mismo de instrumentos musicales, o de sonidos generados electrónicamente por alguna técnica de síntesis –la frecuencia modulada o la modulación en amplitud, por ejemplo. Estos procedimientos compositivos impulsaron la construcción de herramientas apropiadas de análisis y representación del sonido, sin las cuales esta música no podría ser posible. Creo importante el desarrollo en función de la creación, y no al revés. Que las ideas compositivas

fueren la dirección de los desarrollos, no que el compositor deba adaptarse a un sistema que no le pertenece.

A veces, sucede que las herramientas de composición son tan específicas que forman parte misma de la obra. La utilización reiterada del programa no da sino variantes reconocibles del mismo discurso. Otras veces, los programas responden a las ideas del compositor que los diseña, a partir de la formalización de sus propios procedimientos. Si bien existen muchos desarrollos -en el campo de la composición algorítmica, por ejemplo-, su utilización no necesariamente garantiza la formalización de las estructuras de pensamiento de otro individuo. Y en otros casos, el programa es portador de una ideología no compartida, o una imposición estética. Podemos pensar que deberían existir tantos programas como compositores, o incluso obras. En la creación se busca la originalidad, y ésta reside en la búsqueda de un origen de naturaleza interna, que raramente coincide con la ajena. Resulta preferible construir un camino propio, aunque a veces éste nos lleve a redescubrir otros ya transitados.

En nuestro país, y en este medio en particular, no existe aún una conciencia interdisciplinaria que permita un trabajo grupal tras un mismo objetivo. Si alguien posee una idea de cierta originalidad, debe incursionar solitariamente en nuevos territorios para llevarla adelante. Este aspecto puede ser positivo o no, lo cual depende de las dimensiones de lo imaginado. Por otra parte, las condiciones económicas restringen la infraestructura mínima necesaria -algo tan básico como disponer de una cantidad de parlantes iguales para la experimentación, por ejemplo. Todas estas causas hay que considerarlas, y no sólo en función de los resultados finales, ya que pueden comprometer los objetivos iniciales, y atentar contra el entusiasmo.

No obstante, un aspecto interesante del arte contemporáneo local es que los condicionamientos suelen manifestarse metafóricamente, aportando una especial riqueza, no siempre habitual en otras partes del mundo. Las dificultades, contrariamente a lo que pueda

parecer, suelen resultar estimulantes. Por estas latitudes, no existen hoy obligaciones estéticas de ningún tipo, y el artista está solo y sólo comprometido con sí mismo. Este estado le provee un grado de libertad importante, pero a la vez lo ubica en un lugar de encierro, pues lo priva del enriquecimiento que aporta la discusión y el intercambio de ideas.

Francisco Kröpfl, pionero de la música electroacústica en Latinoamérica, funda en 1958 el primer centro de producción musical de la Argentina, en la Universidad de Buenos Aires. La Facultad de Artes y Ciencias Musicales de la U.C.A., y el Centro Latinoamericano de Altos Estudios Musicales, creados por Alberto Ginastera, se convierten en polos de formación de creadores. Ellos, junto a otros, abrieron caminos que no deben ser confundidos ni olvidados, y tanto esfuerzo debe ser finalmente reconocido, pues su negación, sólo nos somete a aquello que rechazamos.

Carmelo Saitta, en una conversación mantenida en torno a estos temas, expresa que el compositor debe eliminar de plano cualquier estructura ajena, que le impida la realización de su propio imaginario poético. Los avances de un mundo desarrollado y globalizado no pueden ser abolidos, ni tampoco desaprovechados, pero resulta imperioso considerar alguna forma de “regionalismo crítico”, que nos permita encontrar nuestras potencialidades auténticas. La actividad artística, desarrollada en consonancia con estas ideas, puede conducir a la resolución de problemas no planteados.

La investigación en torno a la localización espacial del sonido persigue objetivos de lo más variados. Los recursos provienen de la “industria” del cine y la televisión, de las fábricas de videojuegos o electrodomésticos, de las empresas telefónicas y de servicios de realidad virtual, o de la N.A.S.A misma, en su búsqueda de modos de representación de la información en aviones de combate y naves espaciales. Un mercado tan vasto genera cierta ocultación de la información, y ser competitivo en estos términos resulta bastante difícil. Qué queda entonces para nosotros.

La utilización del espacio en la música obedece a razones que exceden a las cuestiones meramente acústicas. Existen motivos poéticos históricos, sociales, culturales, cósmicos o religiosos que determinan la inclusión del espacio en la representación musical. Su uso acompaña a la música desde sus orígenes, y para justificarlo, basta con considerar que toda sensación sonora lleva implícita una determinada cualidad espacial.

La historia acerca innumerables ejemplos sobre el tratamiento de este aspecto del sonido con fines particularmente expresivos. Podríamos remontarnos al canto antifonal y responsorial, y trazar un recorrido por diversos estilos y períodos, para apreciar su evolución. A principios del siglo XVI, por ejemplo, los compositores demuestran un interés particular por la disposición de los recursos vocales en el espacio de representación. Tal es el caso de Ockeghem, o el de Willaert, que con su estilo policoral explota las posibilidades que se presentan en San Marcos, de Venecia, con sus dos órganos enfrentados.

El lugar donde se desarrolla la obra puede ser considerado tan sólo como un marco, o como una instancia capaz de generar la obra misma. Durante el siglo XX han existido muchos intentos destinados a romper la disposición tradicional del concierto, en la cual, instrumentistas y público conforman dos entidades bien separadas. Boulez, por su parte, recurre la antigua idea responsorial en *Répons* (1981-1984), obra compuesta para solistas, un conjunto de cámara, sonidos electrónicos y electrónica en tiempo real. El público se ubica entre los músicos y los parlantes, rodeando al sonido, y siendo rodeado por él. Desde el punto de vista formal, Boulez compara su obra con la forma del Museo Guggenheim, donde el espectador se enfrenta con lo que tiene delante de sí, guarda en la memoria lo que acaba de ver y puede presentir lo que vendrá. Esta forma de espiral en el tiempo es abierta y cerrada al mismo tiempo; si se interrumpe en cierto punto su forma sigue siendo perfecta, pues no carece de nada, pero a la vez, es posible continuarla al infinito.

Otra obra que transforma el espacio de representación es *Il prometeo*, de Luigi Nono, concebida originalmente para San Marcos, pero estrenada finalmente en San Lorenzo en 1984. Aquí, el compositor propone una desvinculación de lo que se ve respecto a lo que se escucha, en una suerte de contrapunto escénico. Los grupos instrumentales, las voces y los sonidos provenientes de los parlantes, interactúan conformando un laberinto sonoro -émulo de la acústica natural veneciana- montado sobre una estructura especialmente diseñada por el arquitecto Renzo Piano.

Estos ejemplos, y tantos otros separados en el tiempo, permiten notar que las características espaciales de la música han sido ampliamente aprovechadas en función de la expresión. Ya sea a través de ideas extramusicales o en vinculación con la estructuración misma de la obra, el espacio cumple una función integradora del discurso. En la actualidad, los medios electroacústicos han aumentado considerablemente las posibilidades de tratamiento del espacio, pues mediante la simulación, es posible multiplicar las fuentes sonoras y distribuirlas en ambientes virtuales, según la voluntad del compositor.

Antes de finalizar con esta introducción, quisiera citar un último ejemplo que considero interesante: *Le noir de l'étoile* (1990) de Gerard Grisey. Esta obra fue escrita para seis percussionistas dispuestos alrededor del público, sonidos grabados y transmisión de señales astronómicas en tiempo real. Grisey dictó clases de composición en Berkeley, entre los años 1982 y 1986. Allí conoció al astrónomo Jo Silk, quien le hizo “escuchar” diversas señales electromagnéticas provenientes de erupciones solares y pulsares, captadas por radiotelescopios -una versión acústica de estas señales puede ser reproducidas a través de un parlante. Los pulsares -residuos cósmicos resultantes de la explosión de supernovas, que giran en el espacio transmitiendo señales periódicas, a la manera de un faro cósmico- producen diferentes alturas o ritmos, de acuerdo a su velocidad de giro, y sólo pueden ser registrados a determinada hora del día. Para esta obra, Grisey utiliza grabaciones del pulsar Véla, que sólo aparece en el hemisferio sur, residuo de la explosión de una supernova que

debió verse claramente hace 12.000 años; y la reproducción en tiempo real del pulsar 0329+54 -generado por una estrella que explotó hace cinco millones de años- cuya señal tarda 7.500 años en llegar a la tierra. Estos sonidos se combinan con la percusión para dar forma a la obra. El estreno tuvo lugar en Bruselas el 16 de Marzo de 1991, a las 17 horas, ya que 0329+54 hace su aparición a las 17:46 en el radiotelescopio de Nançay en Sologne, desde donde se transmitieron las señales al lugar del concierto. Grisey mismo, con sus comentarios acerca de su obra, propone reflexionar acerca de la llegada en directo, al lugar del concierto, de estos relojes cósmicos impasibles, que han franqueado años luz hasta llegar a nosotros. Las señales, se encuentran con una música que no sólo prepara su entrada en escena, sino cuya organización temporal deviene de sus características rotacionales. También sobre el carácter visual, teatral, festivo y didáctico de un suceso que, sin duda, es conmovedor y singular.

Capítulo 1

Factores psicofísicos que determinan la localización de los eventos sonoros

Capítulo 1

Factores psicofísicos que determinan la localización de los eventos sonoros

1.0 Introducción

La psicofísica, al tratar el problema de la localización espacial del sonido, se ha concentrado principalmente en las modificaciones que ocurren en las señales que llegan a nuestros oídos en relación a la dirección y a la distancia del evento sonoro. Los pabellones auditivos, la cabeza, el cuello y el torso actúan de forma combinada, recibiendo los estímulos del campo acústico y transformándolos en su recorrido al oído medio a través de los tímpanos. Cualquier modificación de las características de la señal acústica, relacionada con la posición de la fuente, se convierte en un indicio potencial para la localización.

Suele dividirse a los indicios en dos grupos: *temporales* (cuando su representación en el dominio del tiempo resulta más pertinente para su análisis) y *espectrales* (su representación es más significativa en el dominio de la frecuencia). También es posible clasificarlos como *binaurales* (basados en la comparación de las diferencias percibidas por cada oído) y *monoaurales* (aquellos que percibidos por un único oído, resultan relevantes para la localización).

A partir de los estudios de Lord Rayleigh (1907), se sostuvo por mucho tiempo que derivamos las pistas necesarias para la ubicación de una fuente de las diferencias interaurales de presión sonora y de los tiempos de arribo de las señales acústicas. La diferencia interaural de tiempo parecía aplicable a sonidos cuya frecuencia fuera menor a 1500 Hz, mientras que la diferencia

interaural de intensidad resultaba útil por encima de ese valor. Hoy sabemos que esto es cierto, pero sólo si nos referimos a sonidos puros.

La diferencia interaural de nivel de presión sonora (ILD, por *Interaural Level Difference*) se debe principalmente a la acción separadora de la cabeza, que actúa como una pantalla. Para frecuencias cuya longitud de onda es menor que el diámetro de la cabeza no se produce difracción, sino sombra acústica, registrándose una diferencia de nivel que está en función del ángulo de incidencia de la onda.

La diferencia interaural de tiempo (ITD, por *Interaural Time Difference*) también depende del ángulo de posicionamiento de la fuente, y el pequeño retardo entre las señales que arriban a cada uno de los oídos está asociado a la diferencia entre los caminos recorridos.

Investigaciones posteriores aportaron un indicio monoaural de importancia, expresado a través de las modificaciones espectrales de un evento sonoro en relación a su posición. Estas transformaciones se deben principalmente a la acción del pabellón auditivo, que actúa como un filtro.

Los mecanismos mencionados, junto a otros a analizar, conforman el objeto de estudio de este capítulo. Durante el análisis de las causas que determinan la formación de una imagen espacial del sonido vamos a considerar, en primer término, los fenómenos relacionados con la audición al aire libre. En estas condiciones, las únicas señales que percibimos son las que parten directamente de la fuente y se propagan a través de ondas progresivas, sin encontrar obstáculos a su paso. Es la audición que experimentamos en una cámara anecoica, y en ciertas situaciones naturales donde la reflexión sobre el suelo se encuentra atenuada. Posteriormente, dedicaremos nuestra atención a la percepción espacial en recintos, considerando sus aspectos particulares.

Así como la frecuencia encuentra su correlato sensitivo en la altura, y la intensidad en la sonoridad, denominaremos –siguiendo a Blauert (1997)- “evento sonoro” a una manifestación perteneciente al terreno de los estímulos físicos, y “evento auditivo” a la sensación vinculada a ese estímulo.

Para ubicar a un evento sonoro en el espacio utilizaremos coordenadas esféricas. Definimos la posición de la fuente a través una distancia (r) y dos ángulos, uno medido sobre el plano horizontal –ángulo de azimut (θ), para el que consideramos 0° al frente- y otro que se eleva sobre este plano y mide la pendiente del vector con origen en el centro de la cabeza del oyente, y extremo en la fuente –ángulo de elevación (φ).

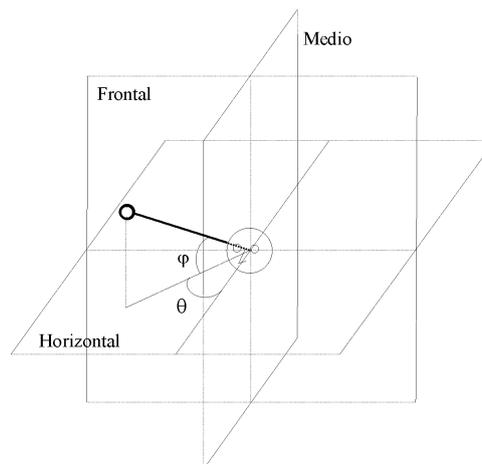


Figura 1.1 Planos relativos al oyente y ángulos de posicionamiento de la fuente

1.1 Funciones de transferencia del oído externo

En esta sección vamos a recurrir a algunos aspectos teóricos de audio digital con el propósito de definir las funciones de transferencia del oído. El lector no familiarizado con este tema, o con los conocimientos matemáticos asociados, puede consultar Moore (1978a y b), Smith (1985) y Moore (1990).

Según mencionamos antes, el oído externo actúa como un filtro, y como tal puede ser definido a partir de su función de transferencia $H(z)$. Esta describe la relación entre la entrada y la salida a través de

$$H(z) = Y(z) / X(z) \quad (1.0)$$

$X(z)$ representa a la transformada z de la señal de entrada e $Y(z)$ a la de salida (figura 1.2).

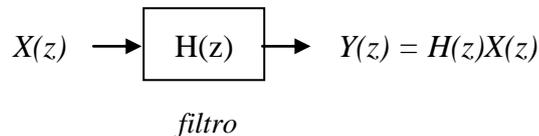


Figura 1.2 Función de transferencia en un sistema lineal

A fin de simplificar esto, podemos definir a la función de transferencia como la relación entre el espectro de Fourier de la señal de salida y el de la señal de entrada. Ella nos informa en qué medida varía la amplitud y la fase de cada una de las componentes de la señal que ingresa al filtro.

Y en términos más precisos, si reemplazamos la variable compleja z por $e^{j\omega}$ podemos calcular la respuesta en frecuencia del filtro $G(\omega)$ a través de $|H(e^{j\omega})|$, y la respuesta en fase $\Theta(\omega)$ por $\angle H(e^{j\omega})$.

Al calcular el corrimiento de fase en función de la frecuencia, producido por el filtro, debemos distinguir entre retardo de fase (τ_f) y retardo de grupo (τ_g).

$$\tau_f = -\frac{\Theta(\omega)}{\omega} \quad (1.1)$$

τ_f es el retardo de fase, y describe cómo varía la fase en función de la frecuencia ($\omega = 2\pi f$). Los gráficos de retardos de fase resultan difíciles de interpretar, porque los valores alternan repetidamente entre $-\pi$ y π para la mayoría de las frecuencias en las que el período es menor a la diferencia temporal. Por esto, conviene a veces expresar los corrimientos de fase a través del retardo de grupo o de envolvente, que indica la proporción de cambio de la fase en función de la frecuencia angular, en particular cuando se trata de bandas estrechas o sonidos modulados en frecuencia. La desviación del retardo de grupo respecto a un valor constante indica el grado de no linealidad de la fase de un filtro, y se obtiene derivando la fase respecto a la frecuencia.

$$\tau_g = -\frac{\partial}{\partial \omega} \Theta(\omega) \quad (1.2)$$

Dado que la multiplicación en el dominio de la frecuencia equivale a la convolución en el dominio del tiempo, podemos también averiguar la señal resultante de un filtro a través de:

$$y(n) = h(n) * x(n) \quad (1.3)$$

donde $h(n)$ es su respuesta a impulso, o sea la señal que egresa del filtro al ingresar un impulso, y el símbolo $*$ representa a la operación de convolución (ver figura 1.3).

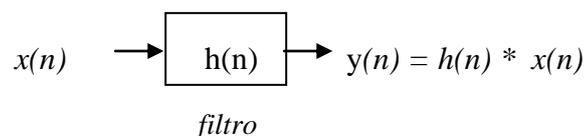


Figura 1.3 Filtro caracterizado por su respuesta a impulso

En la figura 1.4 se observa un gráfico de dos respuestas a impulso, una por cada oído, grabadas con una cabeza artificial, con la fuente ubicada a 30° de azimut y 0° de elevación. El eje horizontal corresponde al tiempo y el vertical a la amplitud. Se aprecian claramente las diferencias entre ambas.

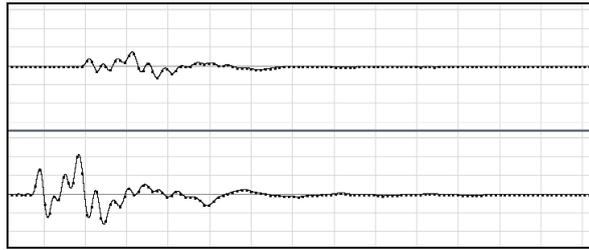


Figura 1.4 Respuestas a impulso registradas con una cabeza artificial para $\theta = 30^\circ$ y $\varphi = 0^\circ$.

Para clarificar estos conceptos, podemos definir al impulso como una señal discreta cuya primera muestra es unitaria y todo el resto nulo, que se representa como $\delta(n)$. El impulso posee una respuesta plana en frecuencia, y su utilización nos permite evaluar el comportamiento de un sistema, tanto en los aspectos temporales como espectrales por su brevedad, y porque posee energía en todas las frecuencias del rango audible.

Si registramos un sonido impulsivo en un determinado ambiente –producido por un disparo, por ejemplo– el recinto responderá a ese estímulo con múltiples reflexiones y una coloración espectral particular. Efectuando la convolución entre la respuesta a impulso de la sala $s(n)$ y otra señal $f(n)$, grabada de un instrumento musical al aire libre por ejemplo, obtendremos un resultado similar al del instrumento ejecutando en ese ambiente $r(n)$. En este caso, hemos evaluado a la sala como si se tratara de un filtro.

La siguiente ecuación (1.4) define a la convolución lineal, que es útil cuando la respuesta a impulso contiene pocas muestras.

$$r(n) = f(n) * s(n) = \sum_{m=0}^n f(m)s(n-m) \quad (1.4)$$

Caso contrario, es más conveniente aplicar la convolución rápida, que se obtiene a partir de la multiplicación de dos espectros (ver figura 1.2). Posteriormente, al tratar la simulación de fuentes aparentes, volveremos sobre este tema.

Es posible pasar de una respuesta a impulso $x(n)$ (una forma de onda representada en el dominio del tiempo) a una función de transferencia $X(k)$ (un espectro representable en el dominio de la frecuencia) utilizando la transformada discreta de Fourier (DFT, por *Fourier Discrete Transform*):

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N} \quad (1.5)$$

N es la cantidad de muestras a analizar, n es el índice que distingue a una muestra de $x(n)$ en particular, k es un entero que se incrementa entre 0 y $N-1$ y representa a las componentes de frecuencia a estimar en el análisis, j es la unidad imaginaria, y según la relación de Euler:

$$e^{-j2\pi kn/N} = \cos 2\pi kn/N - j \sin 2\pi kn/N \quad (1.6)$$

La respuesta a impulso, representada por muestras periódicas en el tiempo, es convertida a un espectro, representado por muestras periódicas en frecuencia. El valor en Hertz que corresponde a cada muestra se obtiene a través de kf_m / N , donde f_m es la frecuencia de muestreo. La aplicación de la DFT da como resultado N posibles componentes en forma de números complejos (a, b) a partir de los cuales obtenemos la amplitud ($\sqrt{a^2 + b^2}$) y la fase ($\tan^{-1} b/a$) de cada componente.

Las funciones de transferencia del oído, reconocidas universalmente a través de la sigla HRTF (por *Head Related Transfer Functions*), dependen de la frecuencia (ω), de la distancia de la fuente (r), del ángulo de azimut (θ) y del ángulo de elevación (φ).

Desde un punto de vista práctico, obtenemos las HRTF de un sujeto o una cabeza artificial ubicando micrófonos en los conductos auditivos y midiendo las respuestas a impulso para diferentes direcciones de la fuente. Estas respuestas, conocidas como HRIR (por *Head Related Impulse Response*) son posteriormente transformadas al dominio de la frecuencia con la DFT, y conforman las HRTF medidas para posiciones particulares de la fuente. La representación gráfica de la magnitud de una HRTF nos permite apreciar cómo varía la amplitud en función de la frecuencia para una posición dada de la fuente, una vez que la señal ingresó en el conducto auditivo. Por otra parte, la comparación de varias HRTF nos da una idea de las transformaciones espectrales que tienen lugar al cambiar la posición del estímulo.

A pesar de la complejidad demostrada por la envolvente espectral de las HRTF, y por la profunda diversidad expuesta en las funciones medidas sobre diferentes sujetos, es posible distinguir ciertas características comunes a todas, según se observa en la figura 1.5 (Kendall y Martens, 1984). Las curvas corresponden a las HRTF medidas sobre dos sujetos, variando el ángulo de elevación entre 0° y 30° en pasos de 10° , con la fuente ubicada a 10 metros.

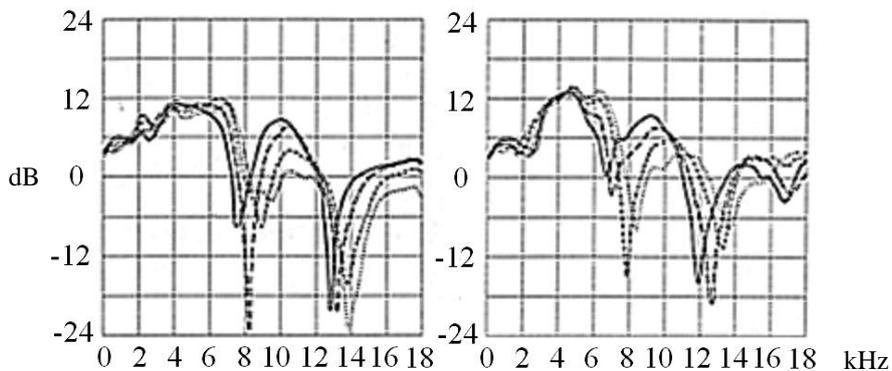


Figura 1.5 HRTF medidas sobre dos sujetos entre 0° y 30° de elevación, en pasos de 10° (Kendall y Martens, 1984)

La línea sólida corresponde a 0°, los guiones largos a 10°, los cortos a 20° y la línea punteada a 30°. Se observan claramente dos importantes depresiones, que aportarían un indicio válido en función del ángulo de elevación de la fuente.

Si $X(\omega)$ es la transformada de Fourier (FT) de la forma de onda que emite la fuente, y si $Y_L(\omega, r, \theta, \varphi)$ es la FT de la señal medida en el tímpano izquierdo, e $Y_R(\omega, r, \theta, \varphi)$ es la FT de la señal medida en el tímpano derecho, entonces las HRTF son:

$$H_L(\omega, r, \theta, \varphi) = Y_L(\omega, r, \theta, \varphi) / X(\omega) \quad (1.7)$$

$$H_R(\omega, r, \theta, \varphi) = Y_R(\omega, r, \theta, \varphi) / X(\omega) \quad (1.8)$$

En los estudios de localización se utilizan tres tipos de funciones de transferencia:

- 1) *Free-field*: relaciona la presión en algún punto del canal auditivo de un sujeto bajo experimentación –preferentemente cerca del tímpano- con la presión sonora sobre un punto coincidente con el centro de la cabeza, cuando el sujeto ya no está presente.
- 2) Monoaural: relaciona la presión medida en el conducto auditivo, para una posición dada, con otra medida en el mismo punto para una posición de referencia (generalmente a 0° de azimut y 0° de elevación).
- 3) Interaural: relaciona las presiones en ambos oídos.

La relación

$$H = H_R / H_L \quad (1.9)$$

es la función de transferencia interaural (ITF, por *Interaural Transfer Function*). Su magnitud, medida en dB, es la diferencia interaural de nivel (ILD), y su fase es la diferencia

interaural de tiempo (ITD). Como se deduce de las ecuaciones 1.7 y 1.8, la ITF es independiente del espectro de la fuente $X(\omega)$

$$H = Y_R(\omega, r, \theta, \varphi) / Y_L(\omega, r, \theta, \varphi) \quad (1.10)$$

Haber eliminado el espectro de la fuente –por aparecer tanto en el numerador como en el denominador de la relación interaural- indicaría que los indicios binaurales no dependen de manera crítica de las características de la fuente, o del conocimiento previo que el oyente posea de ella. Resulta razonable creer que guardamos en nuestra memoria las características aprendidas de nuestras propias funciones de transferencia (Wightman y Kistler, 1997). En la audición practicada con un solo oído, en cambio, como el espectro recibido por el sistema auditivo es el producto entre la HRTF de ese oído y el espectro de la fuente, para recobrar la HRTF que identifica la posición deberíamos conocer de antemano el sonido que esa fuente provoca.

El siguiente gráfico (figura 1.6) representa la magnitud de funciones de transferencia *free-field*, según varios autores. Las mediciones se realizaron a 5 mm de la entrada del canal auditivo a 0° de azimut y 0° de elevación, en cámara anecoica. Se trata de valores promedio obtenidos por Wiener (1947) sobre 6 sujetos (curva *a*), Jahn (1960), también sobre 6 sujetos (curva *b*), Shaw (1966) sobre 10 sujetos (curva *c*) y Blauert sobre 12 sujetos (curva *d*). (Blauert, 1997).

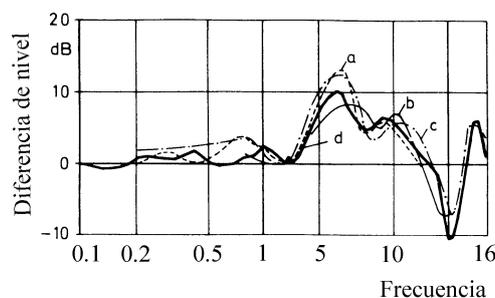


Figura 1.6 Funciones de transferencia *free-field* para $\theta = 0^\circ$ y $\varphi = 0^\circ$ (Blauert, 1997)

La figura 1.7, en cambio, muestra funciones de transferencia monoaurales, medidas sobre distintos ángulos, relativas a 0° de azimut y 0° de elevación. Son valores promedio (25 sujetos) registrados a partir de un impulso producido con un parlante a una distancia de dos metros, en cámara anecoica (Blauert, 1997). La primera columna corresponde a la diferencia de nivel de presión sonora (ΔL), la segunda al retardo de grupo (τ_g), y la tercera al retardo de fase (τ_f), en función de la frecuencia.

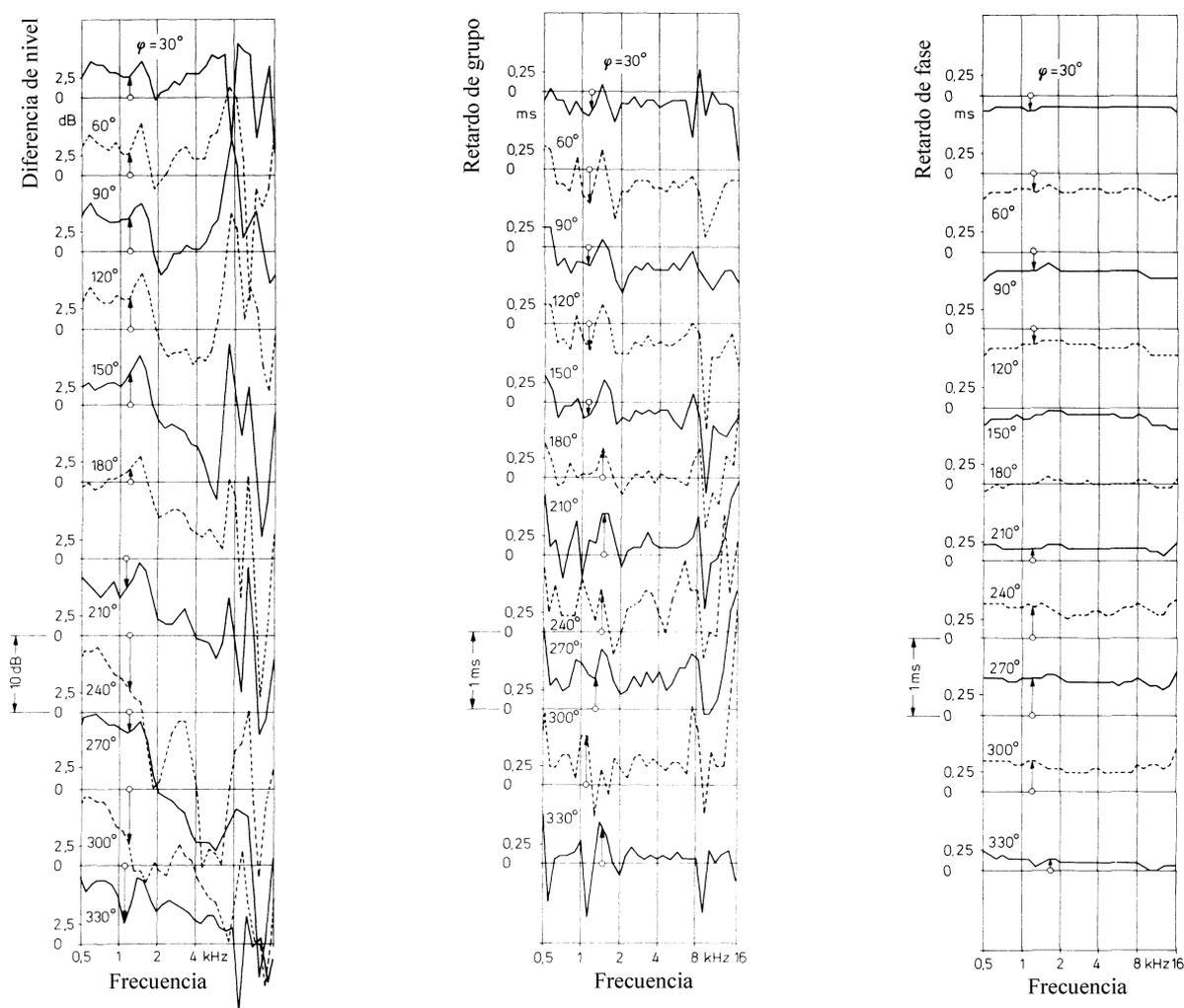


Figura 1.7 Funciones de transferencia monoaurales para una referencia de $\theta = 0^\circ$ y $\varphi = 0^\circ$ (Blauert, 1997)

1.2 Indicios de lateralización y elevación de la fuente

1.2.0 Diferencia interaural de nivel de presión sonora (ILD)

La ILD, producida por los componentes del sistema auditivo externo, es una función sumamente compleja que depende de la dirección de la fuente, de la frecuencia del estímulo, y de la distancia. Para el estudio de las diferentes partes que conforman este sistema suele emplearse un modelo matemático simplificado, en el cual la cabeza, por ejemplo, puede ser vista como una esfera rígida con puntos de detección ubicados a ambos lados.

Diversos autores han contribuido a la creación de modelos análogos. Woodworth y Schlosberg (1954), por ejemplo, propusieron fórmulas para calcular la diferencia de las distancias entre la fuente y cada uno de los oídos (Δs). A partir de ellas es posible obtener diferencias temporales y de nivel de presión sonora. Cuando los rayos dirigidos a los puntos de detección inciden paralelamente -el frente de onda se considera plano pues la fuente se encuentra lejos- la diferencia de recorrido es:

$$\Delta s = \frac{D}{2}(\theta + \sin \theta) \quad (1.11)$$

donde D es la distancia entre los oídos. En otros casos, otras ecuaciones resultan más apropiadas.

Basándose en el modelo de esfera rígida, Shaw (1974) calculó la diferencia de presión sonora entre el punto de detección (supuesto oído izquierdo) y el centro de la esfera (*free field*), generada por difracción para un radio de 87,5 mm. El punto de detección está ubicado a 90°. Los resultados se pueden observar en la figura 1.8.

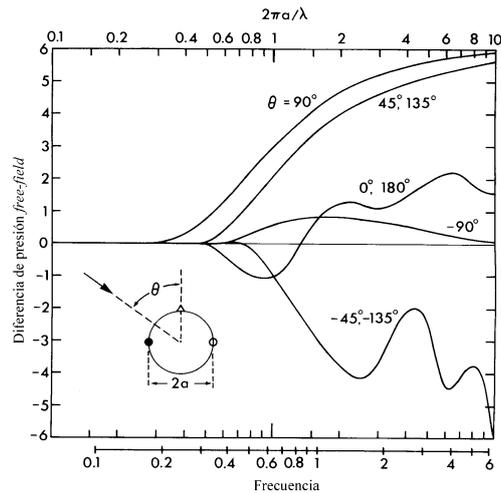


Figura 1.8 Diferencia calculada de nivel de presión sonora entre un punto de una esfera rígida, ubicado a 90° , y *free-field*. (Shaw, 1974)

Cuando el ángulo de incidencia de la onda es de 90° , el nivel de presión sonora crece de forma continua con la frecuencia, hasta 6 dB por encima del valor medido en el centro de la esfera (*free-field*). Para ángulos de 45° y 135° es ligeramente menor. Cuando el oído está en la zona de sombra acústica (por ejemplo a -45° y -135°) los valores caen por debajo del valor sobre el centro, excepto a -90° , donde las ondas difractadas llegan en fase produciendo un incremento de nivel. Si bien se trata de una simplificación, este método ofrece herramientas válidas para la comprensión de estos fenómenos.

Podemos considerar ahora dos puntos, en representación de los dos oídos, y calcular la diferencia de presión sonora y de tiempo entre ambos, es decir, binauralmente. La figura 1.9 muestra cómo varía la diferencia interaural de nivel de presión sonora de acuerdo al ángulo de incidencia. Las curvas fueron calculadas para una esfera de 17,5 cm de diámetro, con puntos de detección ubicados a 100° y 260° respectivamente sobre el plano horizontal (Blauert, 1997).

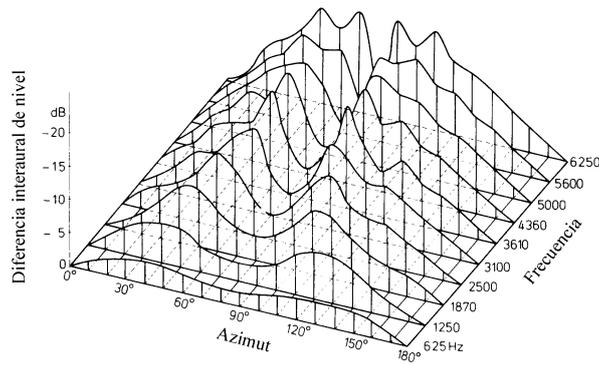


Figura 1.9 Diferencia interaural de nivel de presión sonora, calculada sobre un modelo de esfera rígida. (Blauert, 1997)

La distancia entre la fuente y el sujeto afecta a las diferencias interaurales. Hartley y Fry (1921) ya habían demostrado que la diferencia interaural de nivel sonoro depende en gran medida de la distancia, mientras que la diferencia interaural de tiempo se ve escasamente afectada.

Las funciones de transferencia interaurales, como mencionamos antes, pueden ser medidas ubicando micrófonos adecuados en los conductos auditivos de un sujeto. De este modo es posible registrar las señales que llegan a cada tímpano y evaluar la influencia que el oído externo, la cabeza y el torso tienen sobre ellas. La ILD puede entonces obtenerse a partir de la magnitud del cociente de las HRTF. En la figura 1.10 (Wightman y Kistler, 1997) podemos observar que las ILD son fuertemente dependientes de la frecuencia. El primer gráfico corresponde a una banda de 800 a 1000 Hz, el segundo a otra de 4000 a 5000 Hz, y el último a una banda entre 8 y 10 kHz. Se observa que la ILD es pequeña para bajas frecuencias, y esto se debe a que las dimensiones de la cabeza y el pabellón son pequeñas en relación a las longitudes de onda, dando lugar a la difracción. Las curvas sobre el plano horizontal, en los gráficos, muestran los contornos de ILD constante.

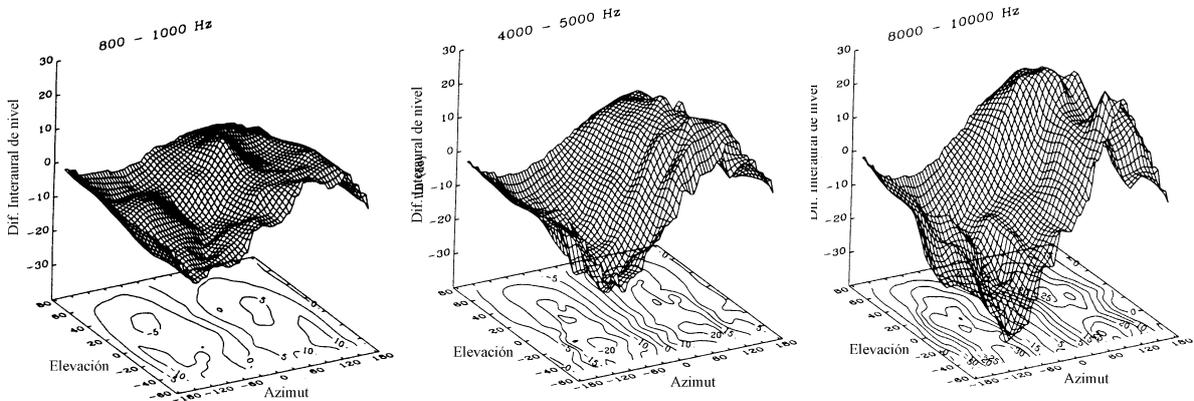


Figura 1.10 Diferencias interaurales de nivel de presión sonora para distintas bandas de frecuencia, medidas sobre un sujeto. (Wightman y Kistler, 1997)

Existen diversos métodos para el registro de las señales que conducen a la construcción de las funciones de transferencia. Con el propósito de minimizar las disparidades encontradas en los diferentes estudios, Shaw (1974) produjo una síntesis de resultados de doce investigaciones, realizadas a lo largo de 40 años en cinco países. Las familias de curvas, presentadas en la figura 1.11, muestran los valores promedio de diferencia de nivel de presión sonora, medidos entre el centro de la cabeza y el tímpano (*free-field*). Se observan las sucesivas transformaciones, modificando el ángulo de azimut a intervalos de 15° , sobre el plano horizontal. Las curvas se han dividido en tres grupos, frontal, lateral y posterior para una mejor discriminación.

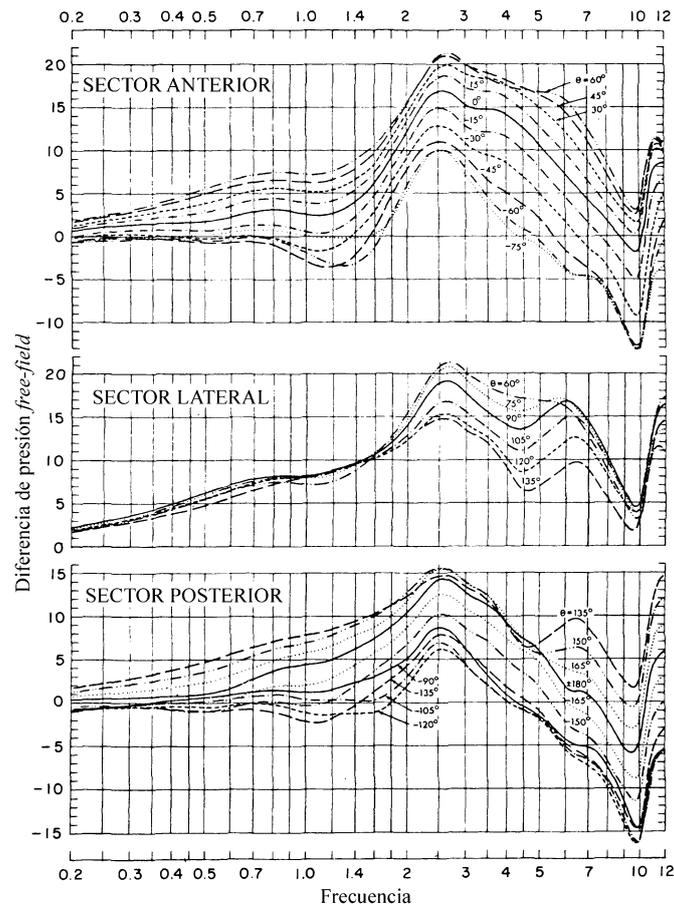


Figura 1.11 Transformaciones *free-field* promedio, sobre el plano horizontal, a intervalos de 15°, obtenidas a partir de una síntesis de doce estudios. (Shaw, 1974)

Algunas conclusiones observadas son las siguientes:

- Una característica común a todas las curvas es el pico alrededor de los 2,6 kHz, producido por la máxima resonancia del oído externo.
- Por debajo de 1 kHz, se aprecia que los valores son mayores a los calculados, posiblemente, por la presencia del cuello y el torso.

- En el sector lateral, lo más característico es la caída de 5 a 10 dB, entre 45° y 135°, sobre la banda de 2,5 a 6 kHz. Este rasgo ha servido de base a teorías de localización en el sector lateral basadas en la coloración espectral.
- La falta de simetría respecto al eje interaural se debe a la difracción en el pabellón auditivo.
- La pequeña depresión entre 1.0 y 1.3 kHz en el sector frontal se atribuye a dos razones. Entre 0° y 60°, debido a la interferencia entre la onda directa y la reflejada en el hombro. Entre -15° a -75°, a la interferencia de las ondas que se difractan, alcanzando el oído en sentido opuesto alrededor de la cabeza.
- La caída en respuesta para todos los ángulos de azimut, en la zona entre 6,5 y 9 kHz, representa una disminución de sensibilidad auditiva del oído en esta banda

La lateralización posee un nivel de definición, entendido como el mínimo cambio en la diferencia interaural de nivel de presión sonora que conduce a un desplazamiento lateral del evento auditivo. Elfner y Perrot (1967) obtuvieron un valor de 2 dB cuando se trata de un sonido puro de 1000 Hz a 60 fones.

El nivel de definición disminuye a medida que aumenta el ángulo de lateralización. Esto ocurre principalmente para sonidos puros de baja frecuencia. Se considera, además, que el nivel de definición depende de la intensidad de las señales. Para intensidades bajas y medias la indefinición decrece, luego permanece constante o crece ligeramente cuando la intensidad sube. También está ligada al aspecto temporal, si las variaciones de posición son rápidas la indefinición aumenta.

1.2.1 Diferencia interaural de tiempo (ITD)

La diferencia interaural de tiempo es uno de los indicios más importantes en la localización espacial. Experiencias realizadas por Wightman y Kistler (1992) revelaron que la ITD podría ser la pista de localización más sólida. Al presentar señales donde la ITD guiaba a la percepción hacia una posición determinada de la fuente, mientras que los otros indicios apuntaban a posiciones diferentes, los juicios se orientaban hacia aquellas ubicaciones señaladas por la ITD, en tanto el estímulo conservara energía a bajas frecuencias. Sin embargo, cuando las frecuencias por debajo de 1500 Hz eran removidas con un filtro pasa altos, la ILD se tornaba dominante, junto con los indicios espectrales monoaurales.

Si realizamos una experiencia de audición de sonidos impulsivos con auriculares, en la cual aplicamos un mínimo retardo a la señal destinada a uno de los oídos, percibimos un desplazamiento del evento auditivo hacia el oído que recibe la señal sin retardar. Aplicando un tiempo de retardo creciente, observamos un desplazamiento lineal cuyo valor máximo ocurre a los 630 μ s. La distancia que recorre el sonido en ese tiempo es de 21 cm, y a este valor se lo conoce históricamente como “constante de Hornbostel-Wertheimer”. Hoy sabemos que no se trata de una constante, dado que depende de diversos parámetros. Desde el punto de vista físico, estos 21 cm corresponden a la diferencia de recorrido entre ambos oídos cuando la fuente está lateralizada 90°.

Una curva típica de desplazamiento lateral del evento auditivo en función de la diferencia interaural de tiempo, para sonidos impulsivos, se ilustra en la figura 1.12, sobre datos de Toole y Sayers (1965).

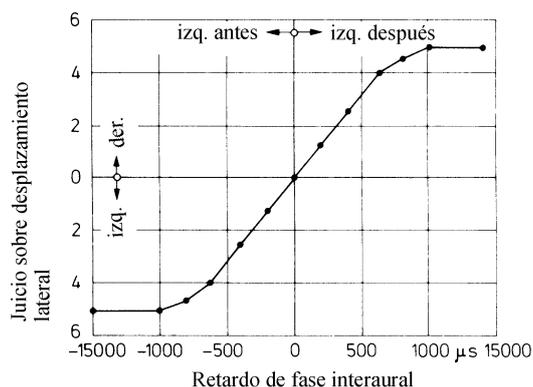


Figura 1.12 Desplazamiento lateral del evento auditivo, medido en función del retardo interaural. (Toole y Sayers, 1965)

El eje vertical corresponde a una escala relativa donde 0 es el centro y 5 el máximo desplazamiento.

El oído no evalúa la diferencia temporal del sonido en su conjunto, sino por bandas. Diversas experiencias con sonidos puros lo confirman. En la figura 1.13 vemos la curva de lateralización para una señal sinusoidal de 600 Hz (Sayers, 1964). Los puntos corresponden a juicios individuales y la línea sólida al valor promedio. El máximo desplazamiento se produce a los 90° de desfase. Cuando la diferencia interaural alcanza un tiempo igual a $T/2$, que equivale a 180°, la mayoría de los oyentes percibe entonces dos eventos auditivos, distribuidos en la posición de cada oído.

Haciendo extensivo este principio a otras frecuencias, el desplazamiento máximo se registra únicamente con sonidos cuyo período, dividido dos, está en el orden de los 630 μs, o sea, para un sonido de unos 800 Hz. Por encima de este valor de frecuencia, el máximo desplazamiento obtenible se torna cada vez menor. Vale decir, que el máximo desplazamiento no ocurre para un retardo de 630 μs, sino para $T/2$, dado que el desplazamiento depende de la frecuencia. Siempre referidos a los sonidos puros, por encima de 1.6 kHz ya no es posible distinguir ningún desplazamiento causado por la diferencia interaural de tiempo.

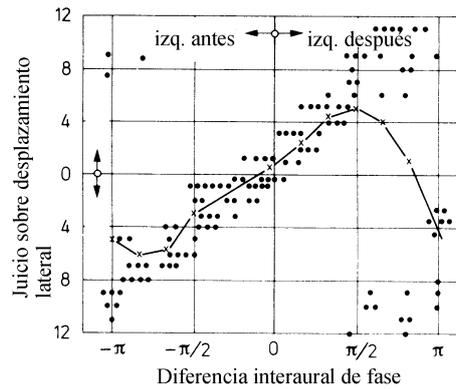


Figura 1.13 Curva de lateralización por retardos para una señal sinusoidal de 600 Hz. (Sayers, 1964)

Scherer (1959) realizó una experiencia a partir de la aplicación de un retardo de $20 \mu\text{s}$ a distintos tipos de señales. En esta prueba midió la habilidad para reconocer una diferencia interaural de tiempo, enfrentando el porcentaje de respuestas correctas con la frecuencia de los eventos sonoros. En la figura 1.14 puede observarse que en el caso de sonidos puros, el número de aciertos decrece rápidamente para frecuencias mayores a 800 Hz (curva 1), y que por encima de 1600 Hz ya no aparecen respuestas correctas. Para sonidos modulados en baja frecuencia ($m = 8 \text{ Hz}$, curva 2) el porcentaje no cae sino después de los 1200 Hz. Por último, para las bandas de ruido con un ancho equivalente a una octava (curva 3), el desplazamiento lateral es reconocible por encima de 1600 Hz.

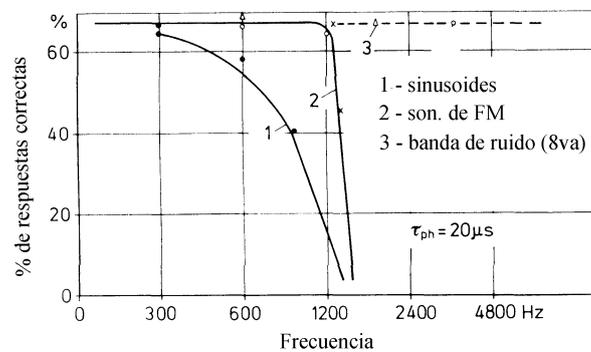


Figura 1.14 Curva de lateralización por retardos para señales distintas. (Scherer, 1959)

Del análisis de esta experiencia se puede inferir que el sistema perceptual recurre no sólo a las diferencias provenientes de la estructura interna de los sonidos (retardos de fase), sino también a cambios temporales relacionados con las envolventes de las señales (retardos de grupo). En otras experiencias comienza a observarse que los corrimientos temporales a nivel de las envolventes, manteniendo en fase a los sonidos, son responsables de los desplazamientos del evento auditivo aún cuando no existen componentes por debajo de los 1600 Hz. Se ha probado, incluso, que el desplazamiento del evento auditivo también se produce cuando los contenidos espectrales difieren, pero el tipo de envolvente se conserva idéntico.

Comparando nuevamente con modelos análogos, es posible calcular los valores de ITD para bajas frecuencias (inferiores a 1.5 ó 2 kHz, con tiempos menores a un período) y altas frecuencias de acuerdo a las siguientes fórmulas (Kuhn, 1987):

$$\text{ITD}_{\text{bf}} = (a/c)(3 \sin \theta) \quad (1.12)$$

$$\text{ITD}_{\text{af}} = (a/c)(\sin \theta + \theta) \quad (1.13)$$

donde a es el radio de la cabeza (aproximadamente 87,5 mm) y c es la velocidad del sonido.

Estas ecuaciones resultan apropiadas porque coinciden con el promedio de las mediciones de diferencia interaural de fase realizadas sobre sonidos graves, y de diferencia interaural de tiempo entre envolventes, para señales agudas de banda ancha (clicks, por ejemplo). Sin embargo, se han encontrado diferencias sustanciales con valores medidos utilizando bandas estrechas (0,22 de octava), en la región entre 4 y 12 kHz, que suelen llegar a los 100 μs (Middlebrooks y Green, 1990). También en la zona de frecuencias intermedias -entre 500 y 2000 Hz- donde aparece una región de transición dada por la dispersión, que hace preciso distinguir entre retardos de fase y de grupo.

Hemos visto, entonces, que existen dos clases de diferencias interaurales de tiempo. El sistema auditivo puede interpretar diferencias temporales tanto a nivel de portadoras como de envolventes. Las diferencias evaluadas sobre las portadoras tienen efecto sólo por debajo de los 1600 Hz, y las de envolvente surten menos efecto a medida que la frecuencia decrece. Las diferencias interaurales de nivel de presión sonora, en cambio, son efectivas en todo el registro audible, siempre dependiendo del tipo de señal.

Desde la neurofisiología practicada en animales ya se han identificado neuronas sensibles a la ILD (Rose, Gross, Geisler y Hind, 1966, por ejemplo), otras a los retardos interaurales de fase en sonidos de frecuencia baja (Goldberg y Brown, 1969), y otras a las envolventes de sonidos complejos de mayor frecuencia (Batra, Kuwada y Stanford, 1989).

Tanto la ITD como la ILD contribuyen conjuntamente a la lateralización del evento auditivo. Algunos estudios se orientaron a la formalización de las relaciones entre ambos indicios, midiéndola a través de un valor denominado “factor de compensación” [$\mu\text{s}/\text{dB}$]. Pero, sin duda, nos encontramos frente a un fenómeno demasiado complejo como para ser descrito en esos términos. Existe una clara dependencia con la sonoridad, que es el principal factor de alteración del factor de compensación, y también con el tipo de señal.

David, Guttman y van Bergeijk (1959), por ejemplo, utilizaron sonidos impulsivos de banda ancha con el propósito de medir las curvas de correspondencia entre la ITD y la ILD. Los resultados se aprecian en la figura 1.15, referidos a distintos niveles de sonoridad.

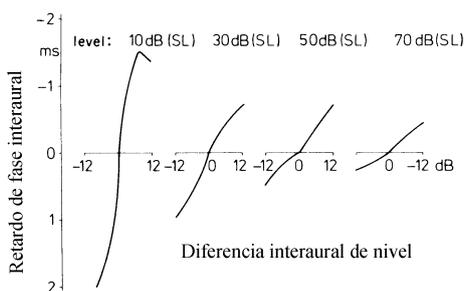


Figura 1.15 Relación entre ITD e ILD para distintas sonoridades. (David, Guttman y van Bergeijk, 1959)

Las evaluaciones interaurales, no obstante, poseen en sí mismas algunas limitaciones. Cuando la fuente se ubica en el plano medio las diferencias son nulas, y en otros casos la diferencia interaural es la misma para varias posiciones. La figura 1.16 muestra el denominado “cono de confusión”, en el cual la resta de las distancias entre cualquier punto ubicado en su superficie, y los puntos que representan a los oídos, es constante.

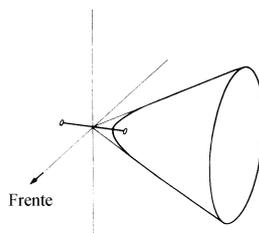


Figura 1.16 Cono de confusión

Estas ambigüedades generan divergencias en la localización, que es preciso resolver a partir de otros mecanismos.

1.2.2 Indicios espectrales monoaurales

Aún las más pequeñas alteraciones de las señales que llegan a los canales auditivos pueden producir notables alteraciones en la imagen espacial. Acústicamente hablando, el pabellón auditivo se comporta como un filtro lineal que afecta fundamentalmente a las altas frecuencias. Produce una distorsión en la señal, en relación al ángulo de incidencia y a la distancia, codificándola con atributos temporales y espectrales. El efecto acústico del pabellón se basa en los fenómenos físicos de reflexión, difracción, dispersión, sombra acústica, interferencia y resonancia, y contribuye tanto a la correcta discriminación entre frente y atrás, como a la detección del grado de elevación de la fuente.

Los pabellones auditivos rodean la entrada a los canales formando un ángulo de 25° a 45° respecto a la superficie de la cabeza, y sus características difieren de un individuo a otro. Por mucho tiempo se consideró que los pabellones auditivos no tenían un rol importante en la audición, considerándolos meros protectores del sistema auditivo. En la actualidad sabemos que cumplen una función determinante en la audición espacial, además de servir a la eliminación del ruido del viento. El canal auditivo, ligeramente curvo y con una longitud promedio de 25 mm, parte de la concavidad central del pabellón y se extiende más allá del tímpano, siempre con un diámetro variado cuyo promedio es de 7-8 mm. El oído externo finaliza en el tímpano, que es ligeramente elíptico y dispuesto a $40-50^\circ$ respecto al eje del canal. La mayor parte de la información que ingresa al oído interno pasa por el tímpano y el oído medio, mientras que otra se transmite a partir de la conducción del hueso temporal, y se le atribuye una importancia secundaria en el estudio de la audición espacial.

Las diferencias interaurales, como indicios, pueden resultar ambiguas en la discriminación vertical sobre el plano medio, y en la resolución de las diferencias que caracterizan a una fuente ubicada al frente o atrás.

Batteau (1967) propuso que las reflexiones producidas en las convoluciones del oído externo eran capaces de brindar indicios adicionales para la localización. Sugirió que los retardos temporales entre el sonido directo y esas reflexiones variaban en función del ángulo de incidencia, transformando así a las señales que tomaban contacto con el tímpano. El proceso era llevado a cabo mediante la evaluación de la delicada estructura temporal resultante, que se observa al apreciar los picos de una respuesta a impulso del oído externo (figura. 1.4). La figura 1.17 muestra un esquema del modelo de Batteau, y las funciones que atribuyó a las componentes del pabellón auditivo.

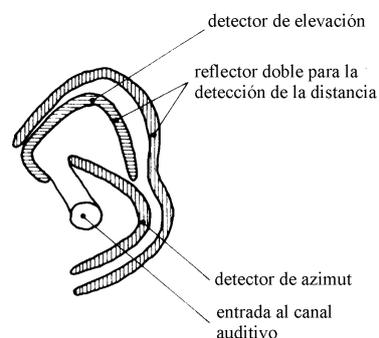


Figura 1.17 Funciones del pabellón auditivo, según Batteau (1967)

Estudios posteriores determinaron que las diferencias temporales percibidas monoauralmente no eran relevantes para la detección de la posición de la fuente. Las HRTF son sumamente breves, alrededor de 2 ms, y la resolución temporal del sistema auditivo se halla en ese orden. Según Green (1971), la estructura temporal de una respuesta a impulso resultaría indetectable.

Partiendo de las observaciones de Batteau, los investigadores se han concentrado más en las modificaciones espectrales derivadas de este fenómeno, que en las transformaciones en el dominio del tiempo. Las convoluciones del oído externo funcionan como resonadores, transformando el espectro del sonido de la fuente en función del ángulo de incidencia.

Shaw y Teranishi (1968) emplearon un modelo artificial y también pabellones naturales, ubicados sobre una placa, para medir la función de transferencia de las señales provenientes

de una fuente alineada con el eje de los oídos, registrándolas con un micrófono. Determinaron con precisión los picos y valles más importantes de la representación de la magnitud de la función de transferencia en relación a la frecuencia. Los resultados más significativos pasaron por la identificación de un número de frecuencias resonantes. Las primeras cinco resonancias ocurren aproximadamente a los 3, 5, 9, 11 y 13 kHz. La primera se atribuye a la acción del canal auditivo y del pabellón como prolongador del canal. La segunda a la concavidad central, y las restantes a ondas estacionarias longitudinales, formadoras de nodos, que dividen a la cavidad central por regiones (figura 1.18).

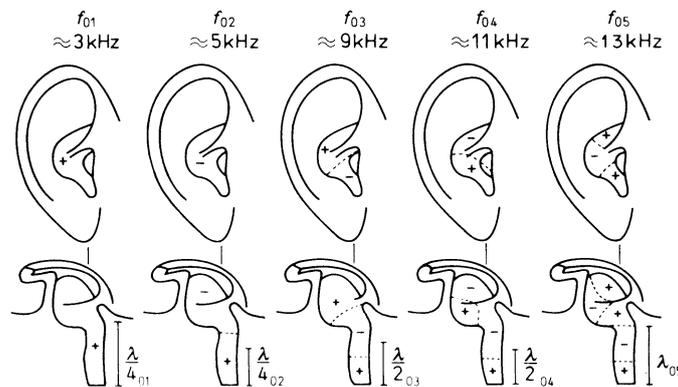


Figura 1.18 Resonancias del pabellón auditivo (Shaw y Tranishi, 1968).

Blauert condujo investigaciones similares en 1967 con el propósito de estudiar las transformaciones de las funciones de transferencia en relación al ángulo de incidencia de la fuente, utilizando un dispositivo similar al empleado por Shaw y Teranishi. Observó que el pico correspondiente a la segunda resonancia mantiene una altura constante cuando la dirección del evento sonoro se encuentra entre 0° y 90° , luego la respuesta decrece en 15-20 dB entre 90° y 110° , y alcanza su menor nivel a los 180° , lo que afirma que las señales provenientes del frente excitan más a esta resonancia que las provenientes de atrás. También demostró que para algunas frecuencias, la dirección influye en el comportamiento de las ondas estacionarias de la cavidad central del pabellón. Confirmó, finalmente, que el pabellón

y el canal auditivo forman un sistema de resonadores acústicos, y que los niveles que adquieren las resonancias del sistema dependen de la dirección y la distancia (Blauert, 1997).

Mientras más ancha es la banda de frecuencias del espectro, mayor es la precisión en la localización. Pero los indicios espectrales percibidos monoauralmente son los menos confiables y parecen depender del conocimiento previo de las características de la fuente. Lo contrario ocurre con la ITD y la ILD, donde el espectro de la fuente se cancela en el cociente de las HRTF, según ya mencionamos. Mediciones realizadas con espectros sistemáticamente alterados han dado como resultado un alto grado de confusión en la localización, tanto en la detección frente-atrás como en el ángulo de elevación, lo que los valida en tanto indicios. Si las diferencias interaurales fueran las únicas responsables de una correcta localización, la alteración del espectro de la fuente no debería influir negativamente. Cuando el sistema perceptual reconoce diferencias interaurales, determina la ubicación de la emisión del estímulo dando un valor más importante a estas diferencias que a los indicios espectrales.

Vemos en la figura 1.19 los resultados de una experiencia realizada por Wightman y Kistler (1997). La primera columna muestra los juicios sobre el espectro plano original, y la segunda, con el espectro modificado luego de cada prueba. El eje vertical representa los ángulos estimados, y el horizontal la dirección real del evento sonoro. Se aprecia el alto grado de confusión frente al espectro modificado, sobre todo en la discriminación de la relación frente-atrás y en la estimación del ángulo de elevación.

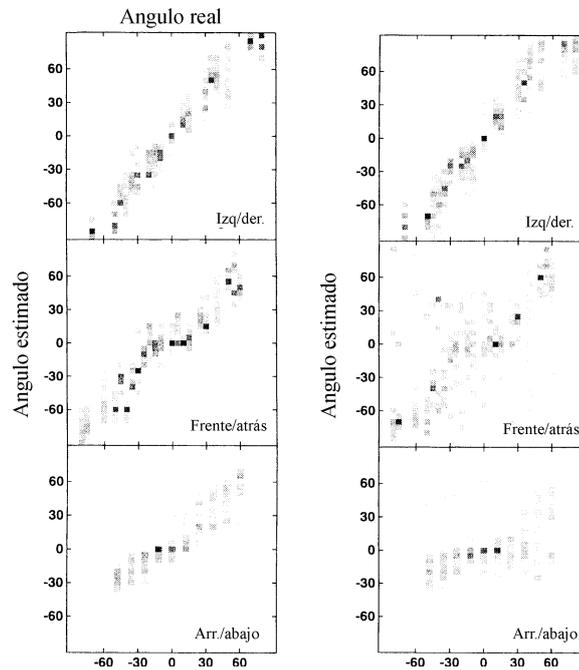


Figura 1.19 Errores en la localización generados por alteraciones sistemáticas del espectro de la fuente (Wightman y Kistler, 1997).

Una forma de aislar la importancia de las modificaciones espectrales como indicio es a través de la audición monoaural, bloqueando uno de los conductos auditivos y eliminando así toda comparación binaural. Diversos estudios se han realizado de este modo (Butler, Humanski y Musicant, 1990, por ejemplo). Middlebrooks midió monoauralmente la precisión de la localización, pero recurriendo a sujetos con deficiencias totales en un oído, sin experiencia en la comparación binaural (Slattery y Middlebrooks, 1994). Los resultados fueron muy precisos, más que en los casos donde se experimentó con sujetos con audición normal, demostrando que las diferencias espectrales son efectivas en la discriminación de la relación frente-atrás, pero también sobre el plano horizontal, aportando información adicional a la lateralización. Vale decir que los indicios espectrales no están confinados al aspecto vertical o la ubicación respecto al plano frontal (Shaw, 1974; Musicant y Butler, 1985).

Las funciones de transferencia del oído para una dirección dada de la fuente han sido denominadas *funciones de transferencia direccionales* (DTF; Middlebrooks y Green, 1990). Las DTF se obtienen a partir de la HRTF para una posición particular, divididas por el promedio de los valores de las HRTF correspondientes a todas las direcciones medidas.

La figura 1.20 (Wightman y Kistler, 1997) muestra las diferencias de magnitud de las DTF registradas en un oyente, transformando el ángulo de elevación mientras el ángulo de lateralización permanece fijo a 90°. Las principales diferencias ocurren entre 5 y 10 kHz, de manera constante a medida que la fuente se eleva. Los indicios aparecen sólo a altas frecuencias, como es de esperar de acuerdo al tamaño del pabellón auditivo. La percepción de la localización es posible cuando existe energía a frecuencias altas, en sujetos sensibles a esta región.

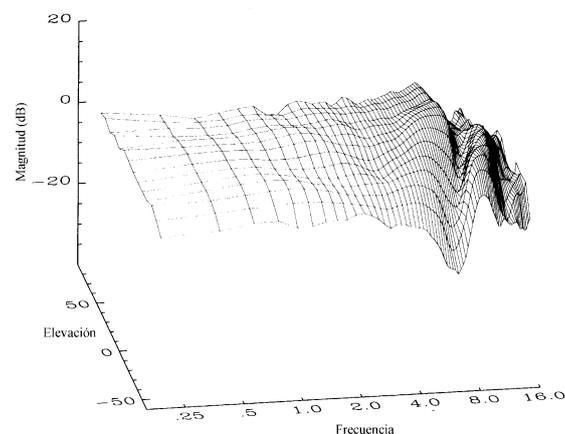


Figura 1.20 DTFs de un sujeto, medidas a 90° de azimut (Wightman y Kistler, 1997).

En el mismo estudio, las DTF de varios sujetos, son comparadas para una misma dirección de la fuente. Es posible observar diferencias que alcanzan los 20 dB, lo cual confirma una vez más que cada sujeto posee HRTF que le son propias.

Middlebrooks (1997) realizó mediciones de las DTF presentando señales de ruido de diferente ancho de banda y duraciones de 150 y 250 ms. La figura 1.21 muestra dos ejemplos, registrados sobre el oído derecho de un mismo sujeto. El gráfico de la izquierda corresponde a ubicaciones sobre el plano horizontal, variando el ángulo de azimut entre 0° y 160° . El de la derecha fue obtenido cambiando el ángulo de elevación entre -60° y 90° (los números sobre las curvas representan a los ángulos) a 0° de azimut. Existen dos curvas por cada ángulo, separadas 10 dB una de la otra. Se observan ciertas particularidades que se presentan sistemáticamente en la envolvente espectral, como el valle, en la figura de la izquierda, cuya frecuencia central se desplaza entre 7 y 9,5 kHz a medida que aumenta el ángulo de azimut.

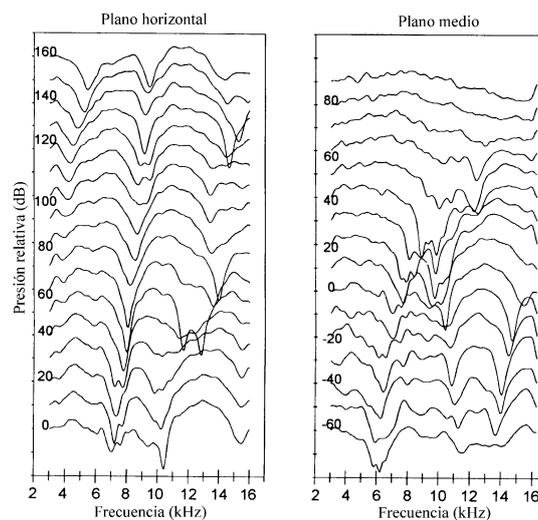


Figura 1.21 DTFs de un sujeto, medidas sobre el plano horizontal y el plano medio (Middlebrooks, 1997).

En otras experiencias (Duda, 1997), destinadas a establecer la dependencia entre la posición de la fuente y la ILD, se partió de respuestas a impulso registradas sobre una cabeza artificial KEMAR, calculando luego las funciones de transferencia interaurales a través del cociente de sus FFT.

Para la realización de este estudio, Duda empleó un sistema de coordenadas basado en conos de azimut constante, que se obtienen manteniendo fijo un ángulo de azimut (entre -90° y $+90^\circ$) y variando el ángulo de elevación entre 0° y 360° (ver figura 1.22).

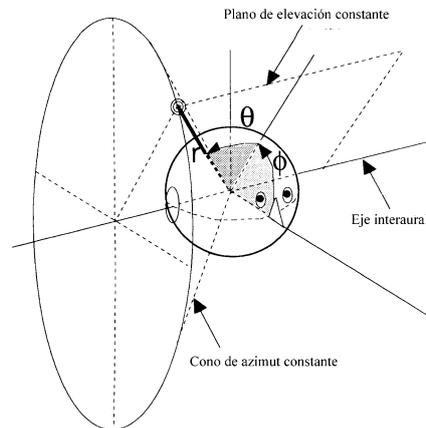


Figura 1.22 Conos de azimut constante (Duda, 1997).

A través de este sistema, la relación frente-atrás depende del ángulo de elevación y no del ángulo de azimut. Searle (1975) y Morimoto y Aokata (1984) ya habían notado que este sistema simplificaba la interpretación de los datos, pues permite agrupar las muestras por conos de igual azimut. Estas figuras, denominadas “conos de confusión” según vimos antes, presentan la misma ITD para todos sus puntos.

Observando la figura 1.23, notamos las diferencias de amplitud y tiempo para una posición $\theta = 30^\circ$ y $\varphi = 0^\circ$. Se aprecia una caída abrupta por debajo de 2 kHz, debida a la acción del parlante. El tweeter no puede reproducir frecuencias por debajo de ese valor, y este efecto de pasa altos limita la información a la zona de registro más significativa. Entre 2 kHz y 6 kHz, la amplitud crece debido a la sombra acústica producida por la cabeza. También vemos un pico prominente a los 7 kHz, producto de la depresión en la respuesta del oído izquierdo, y un valle acentuado a los 11 kHz, debido a la caída en respuesta del oído derecho. Ambos se deben a difracción en el pabellón auditivo.

En el rango entre los 2 y los 18 kHz la diferencia interaural de tiempo fluctúa entre 0 y 1 ms. Esta fluctuación se corresponde con la ubicación de picos y valles del espectro de la diferencia de amplitud.

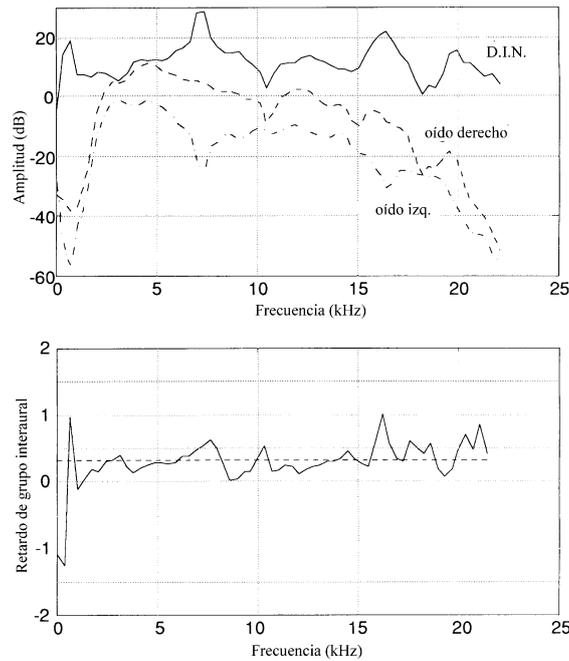


Figura 1.23 Diferencias de amplitud y tiempo medidas

a $\theta = 30^\circ$ y $\phi = 0^\circ$, con cabeza artificial (Duda, 1997).

En relación al sistema de coordenadas elegido para estas experiencias, la figura 1.24 muestra curvas suavizadas de diferencia interaural de intensidad, pertenecientes al cono de 40° de azimut. La elevación varía en incrementos de 10° , comenzando por -90° (curva inferior) y subiendo hasta 230° . Si consideramos una elevación fija, entre aquellas donde la fuente está cerca del piso (cercasas a $\phi = -90^\circ$ ó 230°), observamos que la diferencia de intensidad crece hasta alcanzar su pico máximo alrededor de los 6 kHz. Cuando la fuente está por encima de la cabeza, en cambio, el valor del pico asciende hasta los 10 kHz. La simetría entre las curvas cuyos ángulos provienen del frente (-90° a 90°), y las curvas cuyos ángulos provienen de atrás (90° a 230°), muestra que, si bien la diferencia de intensidad se modifica en función del

ángulo de elevación, no constituye un indicio suficiente para evitar las confusiones en la discriminación frente-atrás. Sin embargo, no existe una simetría exacta alrededor de los 90° . En las frecuencias próximas a 4 kHz, la diferencia de intensidad es mayor para las posiciones del frente que para las de atrás, mientras que lo contrario ocurre alrededor de los 11 kHz. No se sabe si el sistema perceptual examina estas diferencias al evaluar la posición de la fuente, pero los datos están, y parecen resultar suficientes para conformar un indicio.

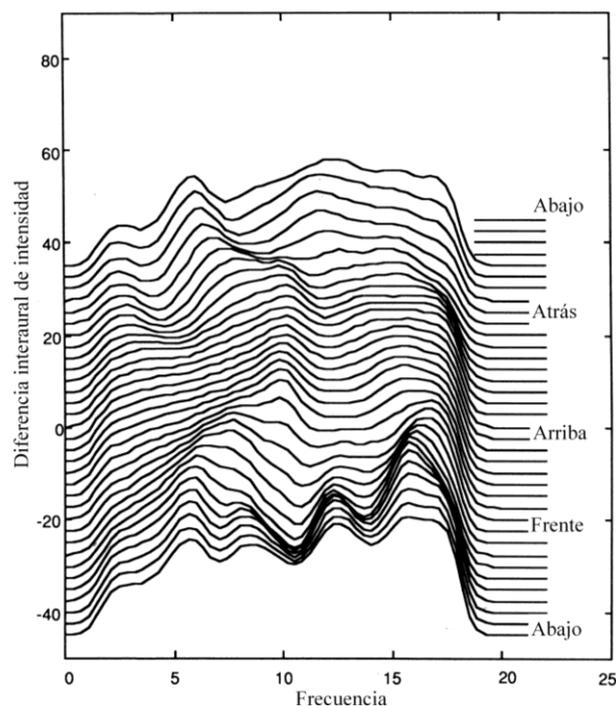


Figura 1.24 Diferencia interaural de intensidad a 40° de azimut y ángulos entre -90° y 230° de elevación, sobre una cabeza artificial (Duda, 1997).

La figura 1.25 muestra la diferencia interaural de intensidad cuando la fuente se mueve sobre el plano horizontal. Se trata de una versión suavizada, donde se aprecia más claramente la conducta de la variación, con características sinusoidales, excepto en la zona de 13 kHz. La falta de simetría marca las diferencias que permitirían discriminar la relación frente-atrás sobre este plano.

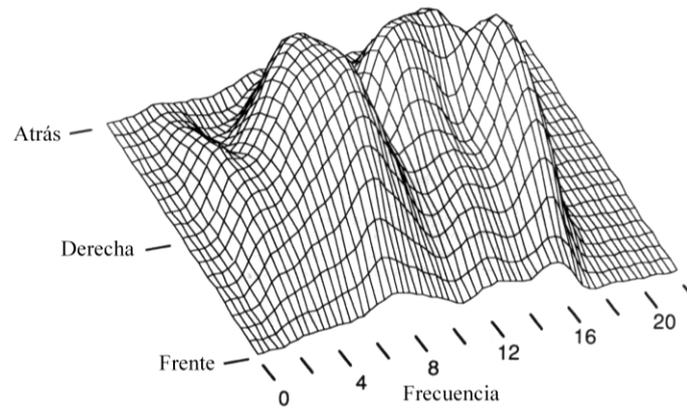


Figura 1.25 Curva suavizada de diferencia interaural de intensidad, para apreciar el contorno casi sinusoidal (Duda, 1997).

Antes de finalizar esta sección vamos a referirnos brevemente a la incidencia del torso en la localización, particularmente para bajas frecuencias. El primero que atribuyó al torso un rol de importancia en la determinación de la elevación en el plano medio fue Gardner (1973), luego de realizar experiencias con sonidos filtrados por encima de 4 kHz. Sus conclusiones, posteriores a la medición de HRTF con y sin presencia de pabellones auriculares y torso, empleando maniqués, determinaron que el pabellón auditivo no ofrece datos por debajo de 3.5 kHz, mientras que el torso aporta información, que podría ser utilizada por el sistema perceptual, entre 700 Hz y esa frecuencia.

Algazi y otros (2000) confirmaron la existencia de indicios de elevación de baja frecuencia, atribuidos a la presencia del torso y la cabeza, no restringidos al plano medio. En sus experiencias realizaron mediciones sobre un maniquí KEMAR, y las compararon con cálculos obtenidos a partir de un modelo geométrico simplificado. Ciertas características particulares que presentan las HRTF por debajo de los 3 kHz, se deben a la reflexión sobre el torso y a la difracción producida por la cabeza.

1.2.3 Correcciones asociadas a los movimientos de la cabeza

Las teorías motoras han estudiado en detalle las relaciones entre la posición del evento auditivo y los cambios de las señales de entrada a los oídos durante los movimientos de la cabeza. Estos movimientos, producidos en el momento de la detección de la posición de la fuente, constituyen un indicio adicional de importancia.

Cuando un oyente mueve su cabeza en la dirección del evento auditivo, buscando la posición probable del evento sonoro, decrece el nivel de indeterminación dado que el mayor grado de precisión se encuentra en la zona frontal. Van Soest (1929) fue el primero en señalar que la resolución de las imprecisiones por falta de indicios concretos, o conflicto entre ellos, se realiza a través de los movimientos de la cabeza.

Vimos que cuando la fuente se ubica sobre el plano medio las diferencias interaurales son prácticamente nulas. En este caso, el movimiento permite generar una diferencia que sirve, en principio, para definir si el movimiento proviene del frente o de atrás. La figura 1.26 puede aclarar este punto. Cuando giramos la cabeza hacia la derecha, si la señal arriba antes al oído derecho, determinamos que la fuente se encuentra atrás.

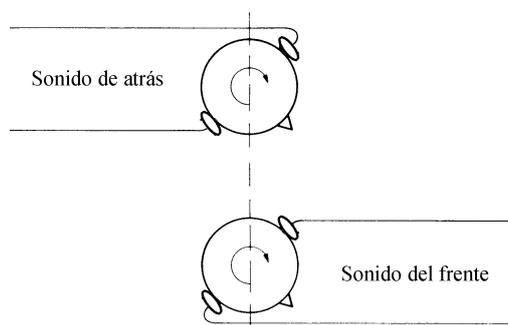


Figura 1.26 Indicio aportado por la rotación de la cabeza.

Van Soest atribuye al sistema perceptual la capacidad de reconocer la polaridad de la ITD como elemento aclaratorio de la posición de la fuente. Para este tipo de evaluación es necesario tener presente el sentido de la rotación realizada, que podría estar indicado visualmente, a través del órgano vestibular del equilibrio, o por el estado de los músculos del cuello.

Los experimentos de Wightman y Kistler también se extendieron a la medición de la incidencia del movimiento de la cabeza en la localización. La posibilidad de orientar la escucha resolvió gran parte de los errores frente-atrás, cometidos previamente con la cabeza inmovilizada (Wightman y Kistler, 1997).

Una anomalía en la detección de la posición de la fuente sonora, especialmente con bandas estrechas, se manifiesta en la percepción del evento en una dirección simétrica respecto al eje que cruza a ambos oídos (Rayleigh, 1877; Fisher y Freedman, 1968, entre otros). Una fuente ubicada en un ángulo de incidencia de 30° sobre el plano horizontal puede ser percibida a 150° (ver figura 1.27). Como vimos antes, el sistema auditivo recurre a las características espectrales para resolver conflictos entre ambas direcciones, pero cuando los sonidos son de banda angosta, esta información es deficiente o nula, dando lugar a una falla en la localización. Una manera natural de resolver esta ambigüedad es mediante el movimiento de la cabeza durante la emisión del sonido. En los casos donde la localización es imprecisa el movimiento exploratorio de la cabeza ayuda a resolver el problema (Blauert, 1997).

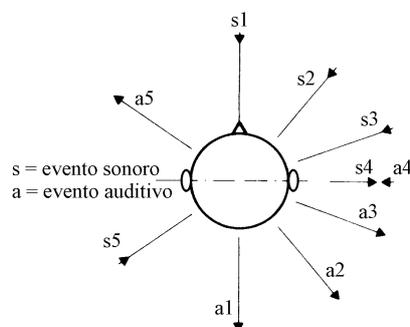


Figura 1.27 Percepción simétrica respecto al eje interaural

1.3 Divergencias en la posición del evento sonoro y del evento auditivo

No siempre la posición del evento sonoro y la del evento auditivo coinciden. Bajo ciertas circunstancias la localización puede ser ambigua, dado que el espacio físico y el auditivo no son necesariamente idénticos. La percepción de la localización puede variar de acuerdo a diversos motivos, entre ellos, por el tipo de señal, por la comparación con eventos anteriores, por la experiencia adquirida en la escucha repetida, por la familiaridad con el sonido de la fuente, por las expectativas del oyente.

1.3.0 Localización indefinida

El concepto de localización indefinida surge de considerar que el espacio auditivo es menos diferenciado que el espacio en el cual los eventos sonoros se manifiestan. El sistema auditivo posee menos resolución espacial que cualquier medición física.

La zona de mayor precisión en la ubicación de la fuente sonora ocurre en la región cercana al punto ubicado justo enfrente del oyente. Es la zona donde un mínimo desplazamiento de la fuente produce una modificación en el evento auditivo. Varios autores han realizado experiencias con el fin de determinar la mínima diferencia perceptible de lateralización en esta zona. El límite inferior corresponde aproximadamente a un ángulo de 1° (Chandler y Grantham, 1992), valor elevado si lo comparamos con la resolución en la visión, que es capaz de distinguir variaciones de 1 minuto de arco.

Los experimentos revelan que la frecuencia central de los eventos sonoros influye decisivamente. El gráfico 1.28 muestra la variación del ángulo mínimo audible cuando la fuente está al frente (MAA, por *Minimum Audible Angle*). El estudio fue realizado con sinusoides (Mills, 1958) sobre tres sujetos, a 50 dB SPL.

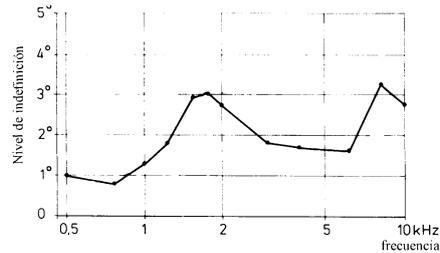


Figura 1.28 Ángulo mínimo audible en función de la frecuencia (Mills, 1958).

Este tipo de experimentos consiste en ubicar un evento sonoro justo al frente y el siguiente lateralizado, interrogando al sujeto si el segundo evento apareció a la derecha o a la izquierda (elección forzada de dos intervalos). La validación del reconocimiento del desplazamiento se realiza cuando las respuestas del 75% de los encuestados concuerdan.

Como dijimos, la menor magnitud de localización borrosa ocurre en la zona del frente. Cuando nos desplazamos a derecha o a izquierda, aumenta. Hacia los lados, este valor se incrementa de tres a diez veces (Stevens y Newman, 1936; Tonning, 1970), y en la parte posterior decrece nuevamente hasta aproximadamente el doble del valor frontal.

El MAA puede variar en función del espectro o la duración de los eventos. Para las señales con un ancho de banda considerable, y en particular las de cierta duración, el error de localización es bastante bajo. Según Roffler y Butter (1968, a, b), y Blauert (1997), el número de aciertos en la determinación de la posición aumenta con el ancho de banda, y también cuando existen componentes por encima de 7 kHz. Para poder discriminar la relación frente-atrás, un ancho de banda entre 2 y 8 kHz es aceptable, pero si aplicamos progresivamente un filtro pasa-bajos, los errores se vuelven cada vez más frecuentes.

El término “persistencia” suele emplearse en relación al tiempo necesario para determinar la ubicación de la fuente. La posición del evento auditivo sólo puede cambiar con una rapidez limitada. Aschoff (1963) realizó un experimento utilizando 18 parlantes dispuestos en una circunferencia. A velocidades bajas de rotación de la señal a través de los parlantes, los

oyentes percibían la trayectoria adecuadamente. A mayor velocidad ya sólo podía percibirse el sonido a la derecha y a la izquierda, cíclicamente. Con un crecimiento aún mayor de la velocidad de rotación, el sonido parecía provenir directamente del centro de la cabeza. Como continuación de este experimento, Plath (1970) demostró que la persistencia es menor cuando la fuente alterna entre derecha e izquierda que cuando lo hace entre frente y atrás.

Los valores de precisión en la localización del ángulo de elevación difieren sustancialmente respecto a lo que sucede en el plano horizontal. Cuando la fuente se encuentra sobre el plano medio, las diferencias interaurales son prácticamente nulas, y rara vez ayudan a la interpretación de la señal. El rango de indeterminación cuando la fuente está ubicada frontalmente es de aproximadamente 17° , si la fuente es una voz hablada que no nos resulta familiar (Blauert, 1997). Si nos encontramos familiarizados con su timbre, el error disminuye a unos 9° al frente, pero es de unos 22° cuando se ubica encima de la cabeza (Damaske y Wagener, 1969). El conocimiento previo del sonido propio de la fuente se convierte en un factor determinante de la correcta localización.

1.3.1 Bandas direccionales

Cuando una señal de banda estrecha proviene del plano medio, o es presentada monoauralmente, y las diferencias interaurales quedan anuladas, la ubicación del evento auditivo depende más de la frecuencia central de la banda que de la ubicación física del evento sonoro. La frecuencia se constituye como un referente espacial, y el primero en estudiar sistemáticamente este fenómeno fue Pratt (1930).

Roffler y Butler (1968) realizaron una experiencia partiendo de sonidos puros de distintas frecuencias, presentados a través de una serie de parlantes ocultos dispuestos verticalmente y enfrentados al oyente. Observaron que la localización no dependía de la posición del parlante

que actuaba como fuente, sino de la frecuencia de las señales sinusoidales. Pueden verse algunos resultados en la figura 1.29.

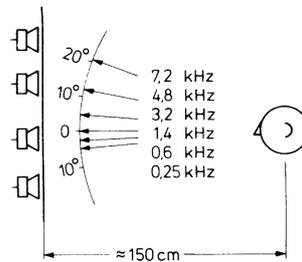


Figura 1.29 Localización del evento auditivo en función de su frecuencia (Roffler y Butler, 1968)

Estos experimentos fueron realizados sobre un gran número de sujetos, algunos de los cuales desconocían el significado del término “altura”. Se probaron distintas intensidades, con la cabeza inmovilizada o libre de realizar movimientos, y todos los resultados fueron similares.

Blauert, también observó que modificando de forma continua la frecuencia de un sonido puro entre 200 Hz y 16 kHz se percibía un movimiento aparente que partía del frente y llegaba hasta atrás, pasando por arriba de la cabeza varias veces. Esto lo condujo a la realización de otras experiencias para determinar el porcentaje estadístico de localización de eventos auditivos al frente, arriba o atrás, con fuentes de ruido de 1/3 de octava. El sonido era reproducido a través de cinco parlantes dispuestos de modo tal que ambos oídos recibieran señales idénticas, ya sea a través de uno o más parlantes. Demostró que los juicios sobre la posición de la fuente dependían de la frecuencia central del sonido reproducido y no de la posición del evento sonoro, ni de la configuración de parlantes usada, ni de su intensidad (ver figura 1.30). A los picos que aparecen en los gráficos los denominó “bandas direccionales”, y la localización de los eventos auditivos en el plano medio podía describirse en función de estas bandas. Si las bandas se centraban en 4 y 6 kHz aparecían al frente, las de 12 kHz por atrás, y aquellas que se centraban en 8 kHz parecían provenir de arriba de la cabeza,

independientemente de la ubicación real de la fuente. Se trata de una ilusión, donde el sonido aparece en un lugar diferente al de la fuente.

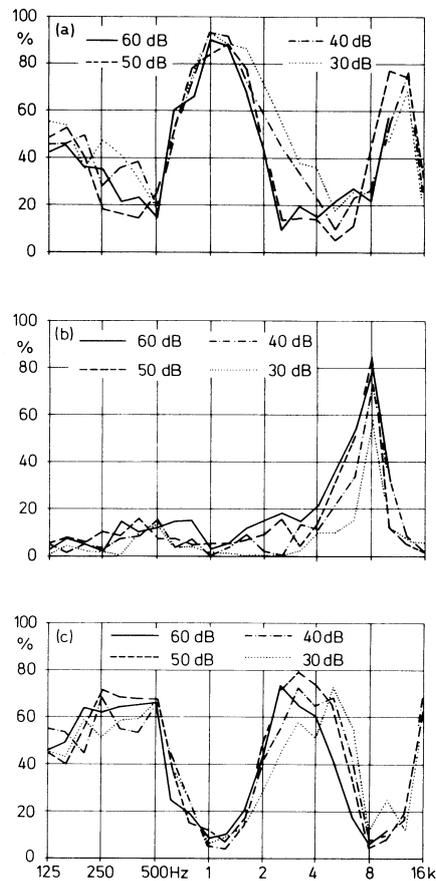


Figura 1.30 Bandas direccionales (Blauert, 1997)

Los tres gráficos corresponden a las respuestas que indicaban que la fuente se encontraba atrás, arriba y al frente, respectivamente.

En una segunda etapa de sus investigaciones realizó mediciones de la función de transferencia del oído, medida al frente y atrás. Al calcular la diferencia de niveles de presión sonora entre esas posiciones observó que el nivel resultante era mayor para algunos rangos de frecuencia cuando el sonido provenía del frente que cuando provenía de atrás, mientras que en otros rangos se verificaba lo contrario.

La figura 1.31 muestra la diferencia de los niveles de presión sonora frente-atrás (curva *v*), y arriba-atrás (curva *o*). El nivel es mayor para 4 y 6 kHz cuando el evento sonoro proviene del frente, y para 8 kHz cuando proviene de arriba.

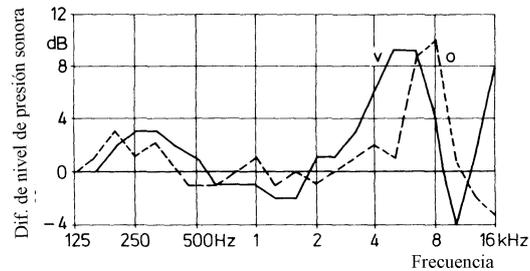


Figura 1.31 Nivel de presión sonora medido en el tímpano con señales al frente (*v*) y arriba (*o*), respecto a la presión medida atrás (Blauert, 1997)

Una banda estrecha de ruido medida en la entrada del canal auditivo produce, para una posición específica, un incremento en el nivel de presión sonora. Si representamos la presión medida por el micrófono, en función de la posición de la fuente, esto se torna evidente. Butler (1987) denominó al punto más saliente de la curva “pico encubierto”, dado que la ubicación donde la banda genera su máximo nivel en el canal auditivo no siempre es evidente cuando el nivel de presión obtenido por el micrófono se grafica en función de la frecuencia, con la posición como parámetro.

Se observa en la figura 1.32, a la izquierda, un gráfico del valor de presión sonora versus el ángulo de azimut, con la frecuencia como parámetro. Para la curva de 11 kHz el pico encubierto ocurre cuando la fuente se encuentra a 80° de azimut. No existe ningún otro punto para esa banda, centrada en 11 kHz, donde la salida del micrófono sea mayor.

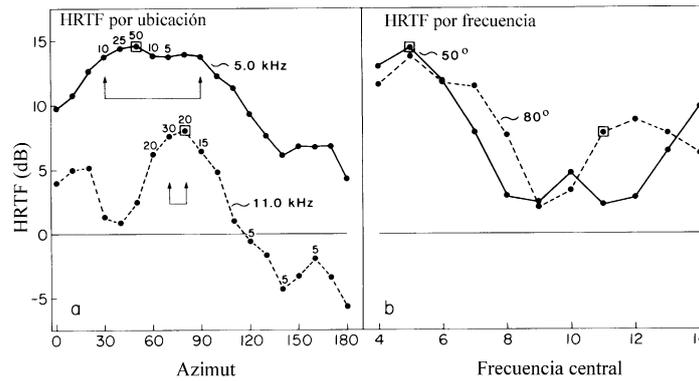


Figura 1.32 Picos encubiertos. Magnitud de la HRTF vs. azimuth y vs. frecuencia, con frecuencia y posición como parámetros, respectivamente (Butler, 1987)

En la figura de la derecha, se representa el nivel de presión sonora en función de la frecuencia central de la banda. El punto que corresponde al los 80° ya no es el más prominente, y se encuentra varios dB por debajo.

Considerando el error asociado con la medición, Rogers y Butler (1992) optaron por denominar a todos los puntos cercanos al punto máximo, cuya diferencia de presión no supere un dB, “área de picos encubiertos”.

Aún se continúa trabajando en esta área, bajo la hipótesis que las bandas direccionales no son sólo una mera curiosidad, sino un indicio probable de localización espacial. La tarea consiste en relacionar las bandas de frecuencia asociadas a una posición aparente en el espacio, con las bandas resonantes medidas físicamente para distintas posiciones (Butler, 1997).

1.3.2 Influencia del grado de coherencia de las señales

Definimos a dos señales como coherentes cuando son idénticas, o bien se diferencian en una o más de las siguientes características:

- Tienen amplitudes diferentes pero la misma forma de onda. Se produce una diferencia de nivel que es independiente de la frecuencia.
- Poseen un retardo de fase que es independiente de la frecuencia
- Se encuentran en oposición de fase

En una experiencia realizada por Chernyak y Dubrovsky (1968) se modificó el grado de coherencia de dos fuentes, presentadas simultáneamente con auriculares, con el propósito de medir la influencia de este parámetro en la representación de los eventos auditivos. Los sujetos fueron provistos con tarjetas, en las que constaba un esquema semicircular representativo del sector frontal superior de la cabeza. Allí debían marcar áreas que describieran la ubicación de los eventos auditivos percibidos. En la figura 1.33 pueden verse los resultados.

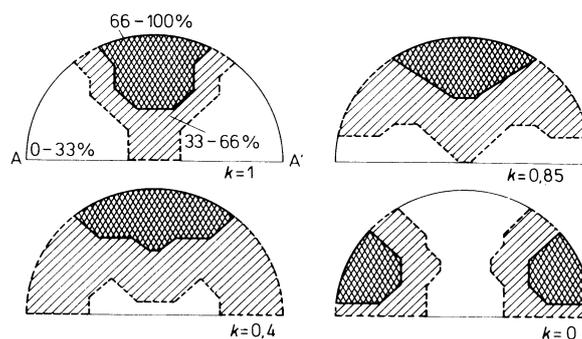


Figura 1.33 Influencia del nivel de coherencia en la ubicación de los eventos auditivos (Chernyak y Dubrovsky, 1968)

Cuando las señales de las fuentes son totalmente coherentes ($k = 1$) aparece un único evento auditivo, en el plano medio, que ocupa una determinada extensión. A medida que el grado de incoherencia aumenta, también crece el área ocupada por el evento. A un mayor aumento, comienzan a percibirse dos eventos diferenciados, uno en cada oído.

La posición de las fuentes influye sobre los resultados. Damaske (1967/8) ubicó dos parlantes ligeramente por encima del plano vertical, irradiando señales de ruido incoherentes. El ángulo de azimut de cada parlante podía ser variado a voluntad. Los diagramas de la figura 1.34 muestran la cantidad de eventos percibidos en relación a las direcciones de los parlantes, así como el área que ocupan.

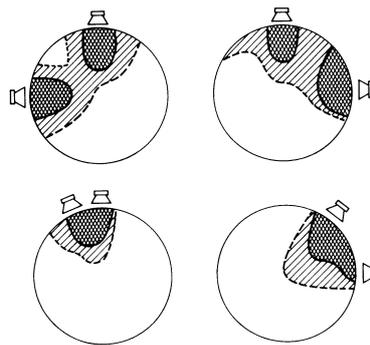


Figura 1.34 Influencia de la posición de dos fuentes incoherentes en la ubicación de los eventos auditivos (Damaske, 1967/68)

1.4 Percepción de la distancia

Con respecto a la determinación de la distancia de una fuente, la familiaridad con el tipo de sonido juega un importante rol en la eficacia de los resultados. La voz hablada a sonoridad normal permite una correspondencia bastante alta entre la posición del evento sonoro y el auditivo. La figura 1.35 muestra resultados sobre la medición de la distancia subjetiva de una voz hablada -susurrando, hablando normalmente y gritando- reproducida por un parlante que se desplaza de 0,9 a 9 metros (Gardner, 1969).

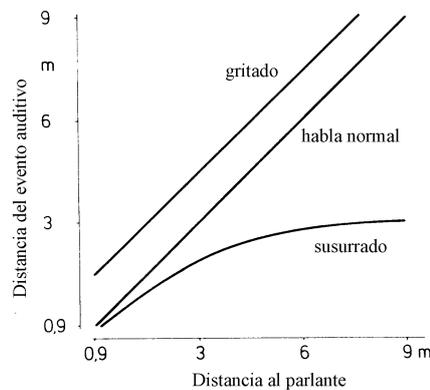


Figura 1.35 Evaluación de la distancia subjetiva en función de la distancia del evento sonoro, para la voz hablada (Gardner, 1969)

Según Blauert, a distancias intermedias (de 3 a 15 metros), el nivel de presión sonora es el atributo que ayuda a determinar la distancia. Al aire libre, el nivel cae 6 dB cada vez que se duplica la distancia.

Una curva típica de la dependencia de la distancia evaluada respecto al nivel de presión sonora se observa en la figura 1.36 (Gardner, 1969). Fue medida con un parlante ubicado a 3 metros y luego a 9 metros. Se deduce claramente que la distancia atribuida al evento auditivo no depende de la distancia del evento sonoro, sino únicamente del nivel de presión sonora que se registra en la posición del oyente.

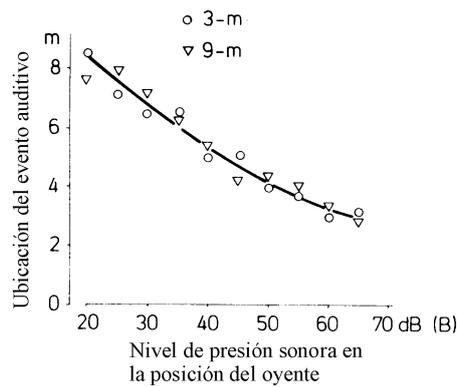


Figura 1.36 Relación entre la distancia estimada y el nivel de la señal, para dos posiciones fijas de 3 y 9 metros (Gardner, 1969)

Para distancias mayores, el nivel de presión sonora continúa ejerciendo su efecto, pero también la acción del aire, que atenúa las altas frecuencias y modifica el espectro de las señales. La figura siguiente (1.37) muestra el nivel de atenuación de una señal que se propaga al aire libre, en función de la frecuencia, las curvas corresponden a una humedad del 40%, con y sin viento. La atenuación $1/r$ propia de la distancia fue removida.

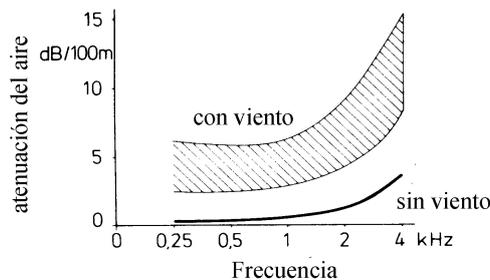


Figura 1.37 Nivel de atenuación del aire, en función de la frecuencia (Blauert, 1997)

Cuando la fuente está cerca, a menos de tres metros, los frentes de onda ya no pueden considerarse planos, y se produce una modificación espectral, debida a la distorsión lineal producida por la cabeza y el oído externo. La función de transferencia se modifica sensiblemente cuando las distancias son pequeñas, y en la actualidad se considera que estas transformaciones constituyen un indicio para la determinación de la distancia.

1.5 Audición en recintos cerrados

La mayor parte de los estudios sobre los efectos ambientales en la localización se centran en la percepción de las características del recinto, en la relación entre los primeros ecos vinculada al efecto de precedencia, y en la estimación de la distancia cuando actúa la reverberación.

Cuando experimentamos la sensación sonora en un ambiente cerrado llega primero a nuestros oídos el sonido directo, que es el que proviene directamente de la fuente. En segundo término las reflexiones de primer orden, que parten de la fuente sonora, chocan con un objeto (la pared, por ejemplo) y alcanzan nuestros oídos, luego las de segundo orden (con dos reflexiones), y así hasta percibir una sensación difusa denominada reverberación. Las primeras reflexiones –en general seis, si consideramos un cuarto vacío con cuatro paredes, techo y piso- contribuyen con la determinación de la posición de la fuente, principalmente si el ataque del sonido es impulsivo. La reverberación, por su parte, nos brinda información sobre las características materiales de la sala, y sobre sus dimensiones.

La cantidad de reflexiones por unidad de tiempo fue establecida por Cremer (1948) a través de la siguiente ecuación:

$$N^{\circ} \text{ de reflexiones / seg} = \frac{4\pi c^3}{V} t^2 \quad (1.14)$$

donde c es la velocidad del sonido, y V el volumen del recinto.

Un ambiente acústicamente ideal es aquel donde la reverberación manifiesta una caída lineal, producto de una absorción uniformemente distribuida en un volumen único, y permanece estática. Esto no ocurre cuando existen áreas adyacentes con volúmenes diferenciados o niveles distintos de absorción.

El tiempo de reverberación (t_{60}) se calcula considerando el intervalo transcurrido entre el comienzo de la reverberación y el momento en que su nivel cae 60 dB. El tiempo de reverberación, cuando el volumen es medido en m^3 es, según Sabine:

$$t_{60} = 0.16 \frac{V}{A} \quad (1.15)$$

donde A representa a la sumatoria de los productos entre los coeficientes de absorción y las superficies reflectoras (el coeficiente de absorción de un piso alfombrado, por ejemplo, es de 0.05 sabins).

El gráfico que sigue (figura 1.38), muestra la distribución de las reflexiones en el tiempo producidas por un impulso dentro de una sala, la altura de las líneas representa su amplitud relativa.

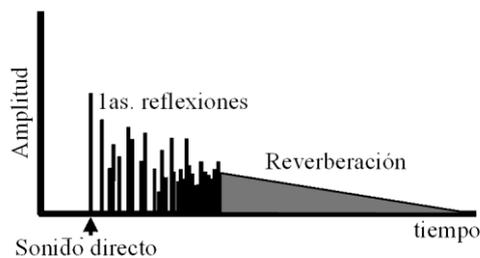


Figura 1.38 Respuesta típica de una sala

Si bien las reflexiones llegan desde diferentes direcciones, atribuimos una posición correcta a la fuente, pues éstas se funden con el sonido directo agregando una sensación espacial más completa y una mayor sonoridad. Este fenómeno de supremacía de la dirección del directo frente a la de los ecos es conocido como “efecto de precedencia” y ha sido objeto de amplio estudio en psicoacústica. Su acción es tan marcada que ocurre aún cuando la intensidad de las reflexiones es hasta 10 dB más pronunciada.

Cuando percibimos dos señales coherentes de forma sucesiva (producidas por dos parlantes, por ejemplo), podemos distinguir alguno de estos tres fenómenos, de acuerdo al intervalo temporal entre ambas:

- a- un único evento auditivo que depende de la posición de los dos eventos sonoros.
- b- un evento auditivo cuya posición es la misma que la de una de las fuentes.
- c- dos eventos auditivos cuyas posiciones son las posiciones de las fuentes.

El primer caso ocurre cuando los niveles y tiempos de arribo de las señales difieren ligeramente. En este caso hablamos de “localización aditiva”, dado que la posición del evento auditivo resulta del promedio de las posiciones de los eventos sonoros.

Si las señales difieren en más de 1 ms, aproximadamente, la posición del evento auditivo es determinada por la señal que llega en primer término. Cremer (1948) denominó a este fenómeno “ley del primer frente de onda”, de gran importancia en la audición en recintos cerrados, pero la primera publicación acerca de este fenómeno corresponde a Henry (1849). En algunos tratados la localización aditiva y la ley del primer frente de onda son consideradas parte del mismo fenómeno, y se lo denomina “efecto de precedencia” (Wallach, Newman y Rosenzweig, 1949).

Si el retardo entre las señales excede cierto valor aparecen dos eventos auditivos sucesivos bien diferenciados, donde el segundo es considerado eco del primero. El límite inferior – umbral del eco- es difícil de determinar pues depende de diversos factores, como el tipo de señal, la dirección de incidencia y la intensidad. Los umbrales de eco más pequeños ocurren con los impulsos, y a medida que la intensidad crece el umbral se torna menor. Los sonidos continuantes, en cambio, poseen un umbral considerablemente mayor.

En el efecto de precedencia, las reflexiones cambian el color del sonido y refuerzan su sonoridad, y si bien no son percibidas como eventos separados del sonido directo, su anulación puede ser reconocida con claridad.

Otro tipo de vínculo se establece entre el efecto de precedencia y la percepción de la voz hablada en ambientes reverberantes, y es conocido como *efecto Haas* (Haas, 1951). Básicamente, se observa que las primeras reflexiones no alteran la comprensión de la comunicación verbal, siempre que arriben en un tiempo no demasiado largo. Es más, ayudan a la percepción al crear un refuerzo de la sonoridad. El interés puesto por Haas en la palabra hablada se debe a que la voz es muy sensible a la reverberación, en contraste con la percepción de la música.

Green (1976) describe efectivamente el efecto de integración entre el sonido directo y las primeras reflexiones. Dice que si parados a 1 metro enfrente de una pared, producimos un sonido impulsivo golpeando dos rocas entre sí, se crea una reflexión que arriba unos 6 ms más tarde, pero que no somos capaces de escuchar. Pero por otra parte, si escuchamos con auriculares dos clicks separados por 6 ms, ambos se perciben claramente. La supresión de la reflexión sucedida en el ambiente no ocurre en la audición con auriculares, y eso es debido al efecto de precedencia.

Un tercer vínculo lo encontramos en lo que suele denominarse “de-reverberación”. La demostración más conocida de este efecto se produce cuando escuchamos una conversación en un determinado ambiente, mientras la grabamos ubicando el micrófono cercano al punto de escucha. Al reproducirla, notamos que el sonido reverberado interfiere con el directo de la voz, dificultando la comprensión, hecho que en la audición natural no ocurría.

Las características temporales del efecto de precedencia en la localización fueron analizadas principalmente por Blauert (1997). Cuando escuchamos dos clicks provenientes de

direcciones distintas, separados por menos de 1 ms, el sistema auditivo percibe al segundo clic integrado con el primero, y determina una localización que es el promedio entre ambas posiciones. Para retardos ubicados entre 1 y 4 ms el efecto de precedencia actúa fuertemente, sin embargo, crea cierta indefinición en la localización, a la vez que aumenta la sensación de espacio. Luego, entre 5 y 10 ms, el efecto comienza a debilitarse: comenzamos a escuchar dos eventos con posiciones diferenciadas.

Las constantes de tiempo del efecto Haas son bastante distintas. Las reflexiones deterioran la comprensión de la voz hablada cuando están retardadas en más de 50 ms. El máximo poder del efecto Haas se encuentra entre los 10 y 20 ms, es decir, cuando el efecto de precedencia desaparece.

Para el efecto de de-reverberación, en cambio, las constantes de tiempo se extienden a varios segundos.

Si los tres vínculos vistos –localización, efecto Haas y de-reverberación- pertenecen al mismo principio de precedencia, sin duda se trata de un proceso sumamente complejo.

Von Békésy (1971) describió un efecto contrario a la ley del primer frente de onda, que observó frente a un retardo de 70 ms. El evento auditivo evaluado para determinar la posición de la fuente era el segundo y no el primero. Blauert (1997) llamó a este efecto “inhibición del sonido directo”. Puede suceder que la reflexión inhiba al sonido directo parcial o totalmente. Para un retardo de 20 ms, el nivel de la reflexión debe ser unos 40 dB más alto para que ocurra el enmascaramiento. Cuando ambos poseen igual nivel, la inhibición nunca ocurre.

1.6 Conclusiones

Ha quedado establecido que los indicios primarios para la determinación de la dirección de una fuente sonora están dados por las diferencias interaurales de tiempo, y de nivel de presión sonora, mientras que los indicios espectrales monoaurales también aportan información, pero fundamentalmente sobre el grado de elevación.

La diferencia interaural de tiempo es extraída por el sistema perceptual, no sólo de las diferencias de fase entre las ondas que llegan a cada oído, sino también de las envolventes dinámicas que las afectan. En el primer caso –retardos de fase- este indicio es efectivo por debajo de 1.6 kHz, mientras que en el segundo –diferencias temporales en el ataque de las envolventes- resulta apropiado para frecuencia mayores, cuyo valor está en función del ancho de banda de las señales.

La diferencia interaural de nivel de presión sonora es un indicio, en general, efectivo en todo el registro, pero el límite inferior de la zona útil varía también con el tipo de señal. Los sonidos graves de banda angosta, por ejemplo, son difíciles de localizar, debido a la difracción. Este fenómeno se produce cuando la longitud de onda es mayor al diámetro de la cabeza, y afecta a las señales –o componentes de una señal- cuyas frecuencias son inferiores a los 1800 Hz.

Los efectos de los pabellones auditivos comienzan a hacerse notorios a partir de los 3 kHz, donde la longitud de onda es comparable a su tamaño. La principal depresión en la curva de respuesta en amplitud del sistema auditivo externo, se manifiesta en la octava entre 6 y 12 kHz.

Los modelos físicos análogos contribuyen a la comprensión de los fenómenos auditivos, y facilitan la formalización de las HRTF. No obstante, es preciso considerar que las funciones

de transferencia varían considerablemente de un sujeto a otro, y que cualquier intento de simulación de los indicios de localización espacial del sonido debe tener en cuenta esta particularidad, derivada de las variantes anatómicas del oído externo.

Las mayores dificultades en la localización surgen cuando la fuente se encuentra en el plano medio -donde las diferencias interaurales son prácticamente nulas- o en la superficie del denominado “cono de confusión” -donde las diferencias son constantes. La falta de indicios concretos genera, además, errores frecuentes en la determinación de posiciones al frente y atrás. Por otra parte, el nivel de definición depende de la posición; las zonas más indefinidas ocurren a los lados, y en la percepción del grado de elevación. Un sistema de simulación destinado a la ubicación precisa de fuentes aparentes, debe proveer información adicional que permita resolver estas ambigüedades.

También es importante contemplar las divergencias que ocurren entre la posición de las imágenes espaciales, y la ubicación real de los estímulos sonoros, pues inducen a errores en la localización que afectan el realismo de determinadas situaciones, propias de la simulación. Sin embargo, estas divergencias, a pesar de sus efectos negativos, pueden servir a la construcción de paradojas psicoacústicas, de utilidad en la composición musical. Fenómenos similares –relacionados con la altura, el timbre o las duraciones- ya han sido usados en la música (Ligeti, entre otros, recurre frecuentemente a ellos). En relación a la audición espacial, el grado de coherencia de las señales reproducidas con parlantes, por ejemplo, contribuye en la creación de imágenes variadas, de interés musical.

Capítulo 2

Simulación de fuentes aparentes en ambientes virtuales

Capítulo 2

Simulación de fuentes aparentes en ambientes virtuales

2.0 Introducción

En este capítulo vamos a analizar algunas técnicas orientadas a la simulación de fuentes ilusorias en ambientes virtuales, y al posicionamiento y desplazamiento del sonido en el espacio.

Dado que la literatura y los desarrollos vinculados con el tema son sumamente extensos, nos concentraremos principalmente en aquellos sistemas que tienen relación directa con la creación de obras electroacústicas o mixtas, y fundamentalmente en sus aspectos prácticos. Los sistemas destinados a otros usos -el cine, aplicaciones multimediales o realidad virtual, por ejemplo- derivan sus principios de lo aquí tratado.

Para la transcripción de algunos de estos modelos de espacialización al ámbito del procesamiento en tiempo real utilizo el entorno gráfico de programación Max-MSP 4.2 (Puckette, 1991). Es un software desarrollado en el IRCAM (*Institute de Recherche et Coordination Acoustique et Musique*), basado en una librería de objetos interconectables, que incluye la posibilidad de desarrollar objetos propios del usuario, escritos en lenguaje de programación "C".

2.1 Simulación a partir de la estereofonía

2.1.0 Fundamentos del sistema estereofónico

El sistema estereofónico convencional reproduce un campo sonoro por medio de dos parlantes ubicados simétricamente al frente, separados en un ángulo de unos 60° . Crea una imagen espacial frontal basada en diferencias de intensidad (ΔI) y retardos temporales (Δt) entre ambos canales. La figura 2.1 ilustra este principio.

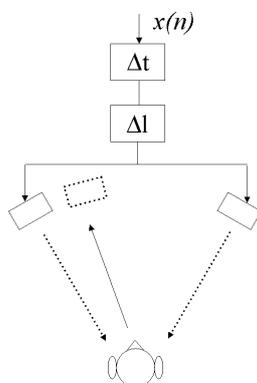


Figura 2.1 Sistema estereofónico. Imagen sonora generada por diferencias de intensidad y retardos.

El sistema es concebido a principios de los años 30, vinculado a técnicas de registro sonoro utilizando dos micrófonos. La técnica MS (*mid-side*), por ejemplo, desarrollada por Blumlein en 1933, utiliza un par coincidente formado por un micrófono cardioide (M) y otro bidireccional a 90° (S) para captar la imagen espacial. Con esta configuración, las señales destinadas a los parlantes (x e y) requieren una decodificación previa, donde se aplica $x = (M + S) / 2$ e $y = (M - S) / 2$.

La técnica francesa ORTF, por otra parte, utiliza dos micrófonos cardioides casi coincidentes, dispuestos a 110° y separados por 17 cm, que asemejan la posición de los oídos en la cabeza (ver figura 2.2).

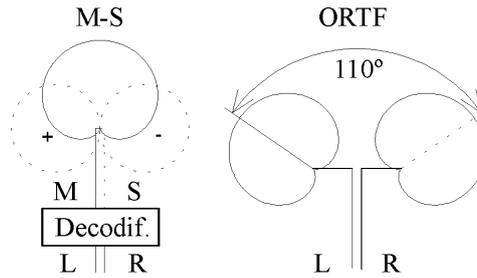


Figura 2.2 Sistema estereofónico. Técnicas microfónicas M-S y ORTF.

Coincidentemente, Fletcher, Steinberg y Snow en los laboratorios de la compañía Bell, desarrollan un sistema de tres canales (Snow, 1953). La demostración tuvo lugar en 1934 con la orquesta de Filadelfia ejecutando de forma remota para una audiencia situada en Washington DC, empleando líneas telefónicas. Esta técnica puede considerarse antecesora del sistema WFS que veremos más adelante.

Si reproducimos una señal continuante que se desplaza de manera equidistante respecto al oyente, la intensidad producida por ambos parlantes debe ser constante en todo momento. Por lo tanto, las ganancias de los canales izquierdo y derecho (g_i , g_d) deben ser tales que se cumpla:

$$g_i^2 + g_d^2 = k \quad (2.0)$$

dado que la amplitud y la intensidad se relacionan a través de:

$$A^2 \propto I \quad (2.1)$$

Las ganancias, el ángulo de posicionamiento de la fuente virtual (θ), y el ángulo que separa a cada parlante del eje que pasa entre ellos (θ_0 , usualmente 30°) se vinculan por la ley del seno:

$$\frac{\sin \theta}{\sin \theta_0} = \frac{g_i - g_d}{g_i + g_d} \quad (2.2)$$

ó bien

$$\frac{g_i}{g_d} = \frac{\sin \theta_0 - \sin \theta}{\sin \theta_0 + \sin \theta} \quad (2.3)$$

Lateralización por control de la intensidad entre canales (*intensity panning*)

Hemos visto que si deseamos generar una trayectoria que parta de la posición de un parlante y alcance al otro, manteniendo la misma distancia respecto al oyente, la ganancia entre ambos canales debe ser tal que dé como resultado una intensidad constante cualquiera sea el ángulo de lateralización. Si A_i es la amplitud del canal izquierdo y A_d la amplitud del canal derecho:

$$A_d^2 + A_i^2 = k \quad (2.6)$$

Podemos comprobar con la ecuación anterior que variando la amplitud linealmente (entre 0 y 1, por ejemplo), cuando la fuente se ubique en el medio ($A_i = A_d = 0,5$) la intensidad caerá a la mitad. Perceptualmente, esto puede ser interpretado como un alejamiento a medida que la fuente se ubica entre los parlantes. Si disponemos los parlantes a 90° , y recurrimos a $\sin^2 \theta + \cos^2 \theta = 1$, obtenemos:

$$\begin{aligned} A_d^2 + A_i^2 &= \sin^2 \theta + \cos^2 \theta \\ A_d &= \sin \theta \quad \text{y} \quad A_i = \cos \theta \end{aligned} \quad (2.7)$$

De este modo, si consideramos 0° a la dirección del parlante izquierdo y 90° a la del derecho, variando el ángulo entre esos valores obtenemos un desplazamiento de izquierda a derecha. Si fijamos 0° en otra posición –al medio, por ejemplo– habrá que realizar los ajustes necesarios.

La disposición de los parlantes en un ángulo de 90° (como en la figura 2.3) si bien no es la más conveniente, la utilizamos para hacer extensible el sistema a cuatro canales (sistema cuadrafónico). Este ángulo de apertura tiende a exagerar la indefinición en la percepción de la imagen espacial cuando la fuente se ubica justo en el medio.

Si disponemos cuatro parlantes formando un cuadrado, podemos calcular las ganancias aplicando la siguiente ecuación (Moore, 1992):

$$g_n = \cos(\theta - \theta_n) \quad \text{si} \quad |\theta - \theta_n| < 90^\circ \quad \text{ó bien} \quad 0 \quad (2.8)$$

donde g_n es la ganancia del parlante n , θ_n es la dirección del parlante n y θ es el ángulo de posicionamiento de la fuente.

Simulación de la ITD

Según vimos en el Capítulo 1, el retardo máximo interaural es de $630 \mu\text{s}$, cuando la fuente se encuentra totalmente lateralizada. Podemos considerar en la simulación a la diferencia interaural de tiempo, aplicando retardos a las señales de los parlantes.

En primer término, debemos conocer la diferencia entre las distancias de la fuente a cada oído en función del ángulo de desplazamiento de la fuente, para obtener luego la diferencia temporal. La formalización más simple es la que corresponde a Hornbostel y Wertheimer (1920), que considera a los oídos como dos puntos en un espacio libre separados por 21 cm, y que no tiene en cuenta la acción de la cabeza. La diferencia en las distancias recorridas por las señales que parten de la fuente y llegan a cada uno de los puntos (Δl) está dada por:

$$\Delta l = d_o \sin \theta \quad (2.9)$$

donde d_o es el diámetro de la cabeza y θ el ángulo de lateralización de la fuente. Luego se realizaron correcciones empíricas de esta fórmula, transformándola en:

$$\Delta l = d_o k \sin \theta \quad (2.10)$$

donde k es un factor equivalente a 1.2 – 1.3.

Posteriormente, fue reemplazada por otras que consideran el recorrido curvo de la señal alrededor de la cabeza (ver figura 2.4). Distinguimos diversos casos, según el modo de incidencia de las ondas (Blauert, 1997):

- dos rayos inciden paralelamente, dado que la fuente está lejos.
- la radiación es desde un punto cercano a la cabeza.
- la radiación es desde un punto cercano, pero uno de los oídos es alcanzado por una trayectoria recta (sin difracción).

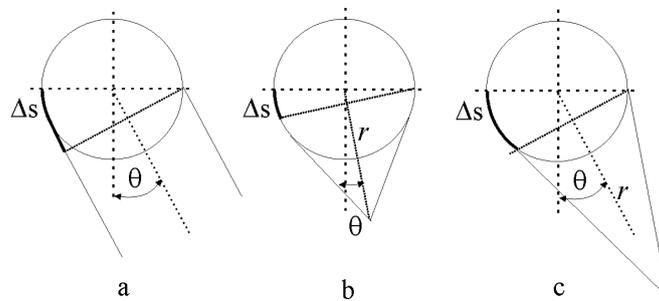


Figura 2.4 Formalización de las diferencias de distancia entre la fuente y cada oído.

Las ecuaciones son, respectivamente:

$$\Delta l = \frac{d_o}{2} (\theta + \sin \theta) \quad (2.11)$$

para $\sin \theta \leq d_o / 2r$, donde r es la distancia entre la fuente y el centro de la cabeza.

$$\Delta l = d_o \cdot \theta \quad (2.12)$$

para $\sin \theta > d_o / 2r$.

$$\Delta l = d_o \left[\left(n + \frac{1}{2} \right) \cos \varepsilon + \frac{1}{2} (\theta + \varepsilon) - \sqrt{n^2 + n + \frac{1}{2} - \left(n + \frac{1}{2} \right) \sin \theta} \right] \quad (2.13)$$

para $n = \frac{r - d_o/2}{d_o}$ y $\varepsilon = \arcsin\left(\frac{d_o}{2r}\right) = \arcsin\left(\frac{1}{1+2n}\right)$.

Ahora es posible calcular la diferencia interaural de tiempo -el intervalo entre el arribo de la señal más cercana y la más alejada- aplicando:

$$\text{DIT} = \Delta l / c \quad (2.14)$$

donde c es la velocidad del sonido.

Moore (1992) propone un valor simplificado, independiente de la distancia, que equivale a:

$$\text{ITD} = 630 \sin \theta \quad (2.15)$$

medido en μs , considerando 0° al frente.

Una línea de retardo variable aplicada al canal izquierdo, por ejemplo, con un tiempo de retardo de $630 \mu\text{s}$, produce una lateralización máxima hacia la derecha. Según vimos en el Capítulo 1, el desplazamiento está en relación con el tipo de señal y su frecuencia media.

Simulación de la distancia

Para simular la distancia (r), aplicamos la ley del cuadrado inverso:

$$A^2 \propto I \propto \frac{1}{r^2} \quad (2.16)$$

y multiplicamos la señal de cada canal por $1/r$. La distancia mínima a considerar debería ser la medida entre el oyente y uno de los parlantes (r_n), y no cero, para evitar el error producido por tal división.

Según Begault (1987) cuando los sonidos producidos por la fuente no resultan familiares, es conveniente aplicar la ley del cubo inverso, y producir una caída de 9 dB en la amplitud al duplicar la distancia.

$$A^2 \propto I \propto \frac{1}{r^3} \quad (2.17)$$

En este caso, multiplicamos la ganancia de cada canal por $1/r^{1.5}$. Con el propósito de permitir evaluar subjetivamente la distancia, este exponente lo ubicamos en una variable modificable por el usuario del programa.

Por otra parte, la absorción en función de la frecuencia que sufre la señal por efecto del aire, aumenta con la distancia. La intensidad en función de la distancia esta expresada en la siguiente ecuación:

$$I(r) = I_0 \frac{e^{-r/\xi}}{r^2} \quad (2.18)$$

donde I_0 es la intensidad de la fuente, $I(r)$ es la intensidad a la distancia r , y ξ es una constante medida en metros que depende de la frecuencia, la temperatura del aire, la humedad y la presión. Para una humedad relativa del 50%, por ejemplo, el valor de ξ para 1 kHz es 0.0013, para 2 kHz es 0.0027, para 3 kHz es 0.006 y para 4 kHz es 0.0207.

Según Moore (1992), la absorción que experimenta una señal para una frecuencia determinada, medida en dB por metro recorrido es:

$$a_a \approx f / 100000 \quad [\text{dB/m}] \quad (2.19)$$

Un sonido de 1.000 Hz, que viaja 100 m, experimenta una atenuación de 1 dB. Otro de 10.000 Hz, para la misma distancia, decrece 10 dB. Vemos así que el aire se comporta como un filtro pasa bajos. Para distancias cortas –menos de 50 metros- esta disminución puede ser inapreciable, pero funciona bien en la realidad para grandes distancias, cuando escuchamos un avión que se aleja, por ejemplo.

Sin embargo, cuando la separación fuente-sujeto es más pequeñas, la aplicación de un filtro pasa bajos en la simulación engaña a la percepción, reforzando la sensación de alejamiento.

Por último, podemos agregar reverberación global, a fin de virtualizar las dimensiones y características del espacio, especificando un tiempo de reverberación y la absorción de agudos en función del tiempo. La relación entre sonido directo y reverberado, por otra parte, agrega mayor realismo a la sensación de distancia. Más adelante retomaremos el tema de la reverberación artificial para verlo con más detalle.

Implementé en 1996 un modelo similar en lenguaje Csound (Vercoe, 1993) con una interfaz gráfica escrita en lenguaje C, para sistema operativo Microsoft Windows. La pantalla principal se muestra en la figura 2.5.

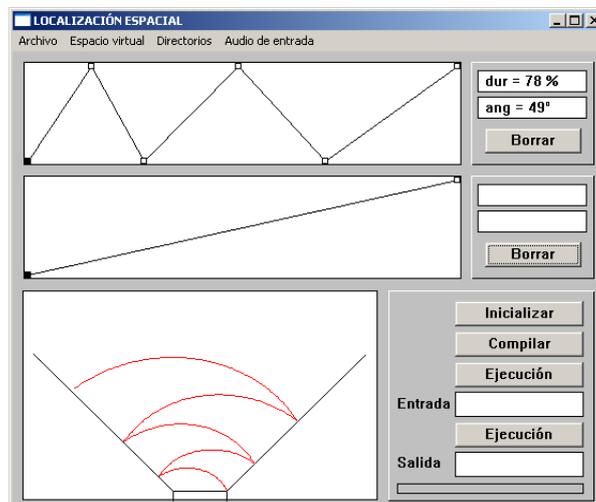


Figura 2.5 Interfaz gráfica del modelo simple implementado en Csound (Cetta, 1996)

2.2 Extensión al sistema cuadrafónico

El sistema cuadrafónico surge en los años 60, y se basa en la reproducción de señales a través de cuatro parlantes que forman un cuadrado, dispuesto horizontalmente. Vamos a analizar las propiedades de este sistema siguiendo el modelo propuesto por Chowning (1971), que implementó en lenguaje Music 10, y con el cual compuso su obra *Turenas* (197?).

2.2.0 Descripción del modelo de Chowning

En la figura 2.6 podemos apreciar un esquema que representa los mecanismos de simulación de este modelo. Al sonido directo se suma reverberación local (diferenciada para cada uno de los canales) y reverberación global (común a todos).

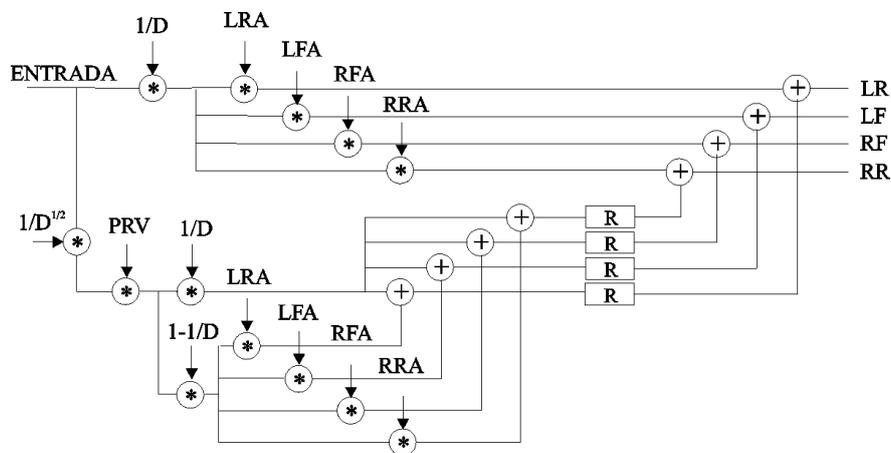


Figura 2.6 Esquema del sistema de localización de J. Chowning.

La amplitud de la señal de audio de entrada es atenuada en función de la distancia ($1/D$), y luego escalada por un coeficiente distinto para cada parlante (LRA, LFA, RFA, RRA) en relación al ángulo de posicionamiento de la fuente (0° - 360°).

La ganancia de cada uno de los cuatro parlantes puede calcularse según vimos antes (ecuación 2.8). Para evitar la realización de cálculos innecesarios (senos o cosenos) se implementa la función que se observa en la figura 2.7 en un oscilador.

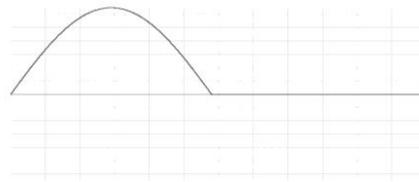


Figura 2.7 Función alojada en un oscilador que evita el cálculo de las ganancias.

Los valores de la función –que varían entre 0 y 1- determinan la amplitud relativa que debe poseer cada parlante para generar un movimiento de 360° . Los cuatro canales leen la misma tabla, pero con fases iniciales diferentes, separados por $1/4$ de ciclo. Vemos que se trata del hemisiclo de una senoide, y que la segunda mitad de la función permanece en el valor 0, que corresponde a la intensidad del parlante cuando la fuente sonora está ubicada del lado opuesto –en el sistema cuadrafónico sólo existen uno o dos canales activos al mismo tiempo-. La función posee 512 muestras, por lo que resulta necesaria un escalamiento del valor del ángulo de posición (0 a 359°) a un número de muestra dentro de la tabla que la aloja (0 a 511).

La parte inferior del esquema del modelo (figura 2.3) muestra la rama de tratamiento de la reverberación. La señal entrante es atenuada en función de la distancia, pero ahora por la raíz cuadrada de la inversa de la distancia, y posteriormente por un valor empírico (PRV) que representa al porcentaje de reverberación. La rama de reverberación es luego escalada por $1/D$ a nivel global, mientras que a nivel local es atenuada por $1-1/D$, y por los coeficientes en función del ángulo de posicionamiento de la fuente. El sistema utiliza cuatro unidades de reverberación decorrelacionadas, con el objeto de hacer más realista la simulación.

Otra característica saliente del modelo de Chowning es la simulación del efecto Doppler. Cuando una fuente está en movimiento, percibimos una variación en la altura del sonido, que depende de su velocidad. Cuando la fuente avanza en el mismo sentido que el frente de onda, la longitud de onda (λ) se comprime por efecto de la superposición de los movimientos de igual sentido, y la altura crece. Cuando la fuente se aleja ocurre lo contrario, ya que la longitud de onda y la frecuencia se vinculan por:

$$f = c / \lambda \quad (2.20)$$

donde f es la frecuencia y c es la velocidad del sonido.

El efecto Doppler se produce naturalmente al emplear una línea de retardo variable. La modificación dinámica del tiempo de retardo trae aparejado un cambio en la altura del sonido, que es equivalente a la producida por este efecto. Si conocemos la distancia entre la fuente y el oyente (r), podemos calcular el tiempo que tarda en llegar el sonido (t) por:

$$t = c / r \quad (2.21)$$

Ese tiempo de llegada es el que se aplica como tiempo de retardo, y al variar debido a que la fuente se desplaza, genera un cambio en la frecuencia de la señal.

2.2.1 Implementación en tiempo real

A continuación, vamos a examinar una implementación en Max-MSP que realicé a partir del modelo antes explicado.

En la figura 2.8 se muestra la unidad, denominada *Turenas* –por el título de la obra de Chowning–, donde se encapsuló el modelo. Los parámetros que recibe son –de izquierda a

La figura 2.10 muestra en detalle al objeto *c_amp*, donde se realizan las operaciones destinadas a obtener cada coeficiente de amplitud en función del ángulo. Según mencionamos antes, los cuatro canales leen la misma tabla (con la función *peek*) pero con diferencias de fase de 1/4 de ciclo.

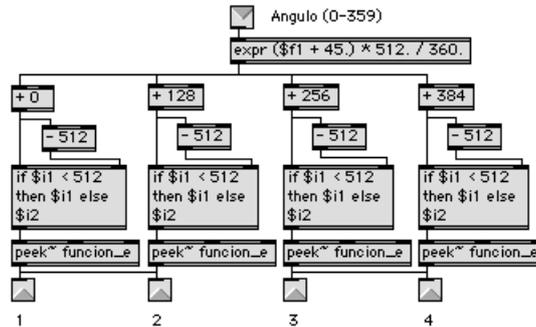


Figura 2.10 Unidad de lectura de los coeficientes de amplitud del objeto *Turenas* (Cetta, 2003a).

La línea de retardo variable, que simula el efecto Doppler reside en un objeto aparte, que puede agregarse.

2.3 Ambientes virtuales

Mencionamos, en el Capítulo 1, que la respuesta a impulso de un ambiente puede dividirse en tres partes: sonido directo, primeras reflexiones, y reverberación. El sonido directo es el que representa más fielmente a la fuente, dado que la única transformación que sufre es el efecto de filtrado producido por la acción atmosférica. En esta sección veremos cómo simular las primeras reflexiones y la reverberación.

2.3.0 Simulación de las primeras reflexiones

Existen diversos métodos geométricos que permiten calcular el comportamiento de las reflexiones en un recinto cerrado. El método de trazado de rayos (Krockstadt, 1968), por ejemplo, consiste básicamente en la generación sistemática de recorridos, y en la evaluación de aquellos que alcanzan la posición del oyente. Emplea algoritmos basados en la probabilidad y en cadenas de Markov. El trazado de haces de rayos (Dadoun y otros, 1985), por otra parte, reemplaza a los rayos individuales por conjuntos que forman haces piramidales, y no sólo es de utilidad en el diseño acústico de salas, sino en proyectos de iluminación, propagación de ondas de radio y determinación de la visibilidad en gráficos en tres dimensiones.

Dado el tipo de aplicación que perseguimos, donde no existe una referencia visual que guíe o ponga en conflicto la percepción acústica, podemos simplificar el diseño del ambiente virtual eliminando formas geométricas complejas, objetos en su interior, bordes que produzcan difracción y otras consideraciones propias del diseño acústico de salas reales de concierto. De este modo, el problema se reduce al mero cálculo de las primeras reflexiones, el tiempo de reverberación, y la absorción de las paredes, el techo y el piso dentro de una caja de determinadas dimensiones.

En nuestro caso, el método de las fuentes imaginarias (Allen y Berkley, 1979) se adapta perfectamente al propósito que perseguimos. Consiste en la determinación de las reflexiones especulares de la fuente, a partir del rebatimiento de la planta sobre ejes coincidentes con las paredes del recinto. Para simplificar, consideremos el problema en dos dimensiones. Uniendo la posición de la fuente rebatida (fuente fantasma ubicada fuera del ambiente físico) con la posición del oyente sin rebatir, obtenemos el punto de intersección con la pared, que coincide con el punto de reflexión del rayo entre fuente y sujeto. La figura 2.11 ilustra este principio, para las reflexiones de primer y segundo orden.

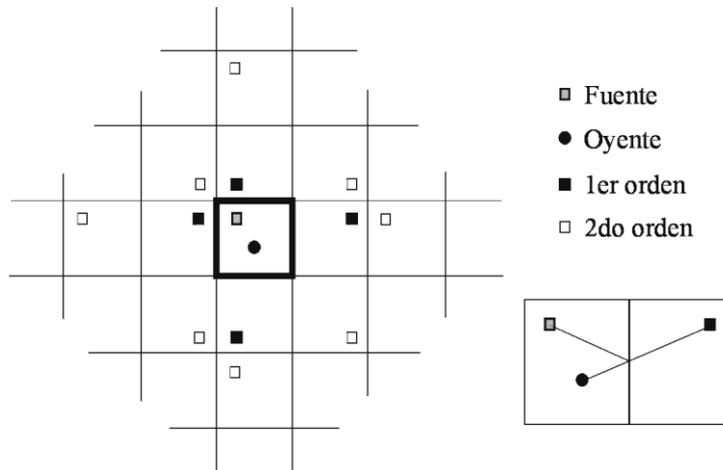


Figura 2.11 Método geométrico de las fuentes imaginarias para el cálculo de las reflexiones en un recinto.

En consecuencia, es posible simular las reflexiones generando fuentes virtuales en las posiciones que el método indica. Dadas las dimensiones del ambiente, la ubicación del oyente y la posición de la fuente, obtenemos las direcciones y distancias de las fuentes virtuales a través de cálculos elementales.

Una vez obtenidas las distancias, simulamos las reflexiones empleando líneas de retardo.

La línea de retardo produce un retraso en el tiempo, expresado en cantidad de muestras (D) de la señal. El tiempo de retardo equivale a DT , donde T es el período de muestreo (tiempo que dura una muestra).

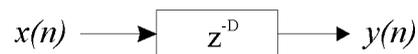


Figura 2.12 Línea de retardo.

$$y(n) = x(n - D) \quad (2.22)$$

para $n = 0, 1, 2, \dots$ y $x(n) = 0$ si $n < 0$

La línea de retardo puede presentar un control de ganancia a la salida, que resulta útil cuando se desea simular un eco. Conociendo la distancia recorrida por la onda reflejada (r), el retardo en cantidad de muestras (D) está dado por $D = r/cT$, y la atenuación del eco por $1/r$. El control de ganancia, además, puede ser reemplazado por un filtro pasabajos, de manera que la atenuación sea proporcional a la frecuencia de la señal de entrada.

Para generar múltiples ecos se utilizan líneas de retardo con derivación conectadas en serie (figura 2.13), con tiempos de retardo distintos (*multitap delay*).

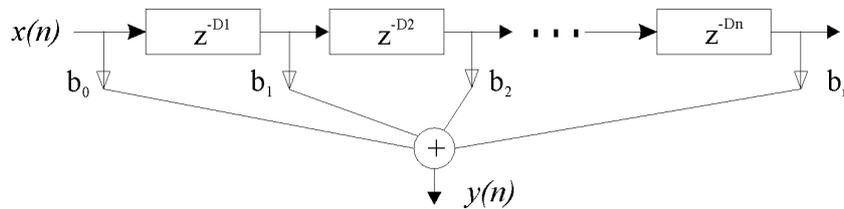


Figura 2.13 Simulación de las primeras reflexiones con líneas de retardo en serie.

Se aprecia la similitud con el filtro causal de respuesta finita al impulso (FIR) en su forma general, donde cada línea produce un retardo de una muestra.

Dado que también conocemos la dirección de las fuentes virtuales que simulan las reflexiones, podemos espacializarlas, utilizando nuevas instancias del método de localización empleado.

Si bien la reflexión trae asociados fenómenos de difusión, el método de las fuentes imaginarias provee una aproximación eficaz, tanto en dirección como en tiempos de arribo de los primeros ecos.

2.3.1 Simulación de la reverberación

Las salidas de la cadena de retardos –especialmente las que corresponden a las reflexiones de segundo orden - ingresan a la unidad de reverberación para multiplicarse y decorrelacionarse a través de redes de filtros. Vamos a comenzar esta sección con una enumeración de los tipos de filtro que constituyen un reverberador, para mostrar luego algunas configuraciones ya establecidas.

Filtros *comb*

Los filtros *comb* toman su nombre de la forma en que se manifiesta su respuesta en frecuencia, cuyos picos sugieren la forma de un peine.

Distinguimos básicamente dos tipos de filtro *comb*. El primero, *feedforward comb*, está compuesto por una línea de retardo cuya salida se adiciona a la señal entrante, y produce un efecto de único eco, atenuado por el valor b_D (ver figura 2.14).

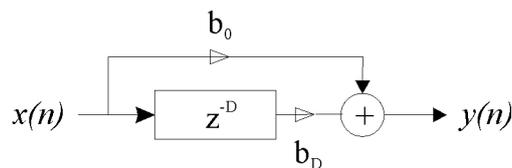


Figura2.14 Filtro *feedforward comb*.

Su ecuación es

$$y(n) = b_0 x(n) + b_D x(n - D), \quad (2.23)$$

y su función de transferencia $H(z) = Y(z)/X(z)$, donde

$$H(z) = b_0 + b_D z^{-D}, \quad (2.24)$$

de la cual obtenemos la respuesta en frecuencia entre $-\pi \leq \omega \leq \pi$ con

$$G(\omega) = |H(e^{j\omega})| = |b_0 + b_D e^{-j\omega D}| \quad (2.25)$$

El segundo, es el filtro *feedback comb*, formado por una línea de retardo cuya salida, retardada en D muestras, se suma a la misma señal entrante, y produce un efecto de múltiples ecos periódicos, atenuados en cada reingreso por el valor $-a_D$ (figura 2.15).

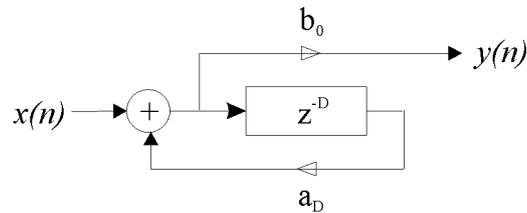


Figura 2.15 Filtro *feedback comb*.

Su ecuación es

$$y(n) = b_0 x(n) - a_D y(n - D) \quad (2.26)$$

y para que el filtro sea estable, debe cumplirse $a_D < 1$. En el caso particular donde $b_0 = 1$ y $-a_D = g$, la ecuación puede escribirse como

$$y(n) = x(n) + g y(n - D) \quad (2.27)$$

Su función de transferencia $H(z) = Y(z)/X(z)$ es

$$H(z) = \frac{1}{1 - g z^{-D}}, \quad (2.28)$$

de la cual obtenemos la respuesta en frecuencia para $-\pi \leq \omega \leq \pi$ con

$$G(\omega) = |H(e^{j\omega})| = \frac{1}{|1 - g e^{-j\omega D}|} \quad (2.29)$$

Cuando $g > 0$, los picos de la respuesta en frecuencia se ubican de acuerdo a:

$$\omega_k = 2\pi k / D \quad (2.30)$$

donde k es un entero que varía entre 0 y $D-1$. Cuando $g < 0$, los picos ocurren entre medio de esos valores.

Filtro *allpass*

Como su nombre lo indica, el filtro *allpass* deja pasar la totalidad de la señal, pero altera la fase de las componentes. Se trata de un filtro cuya ganancia es unitaria para todas las frecuencias, y puede considerarse como una combinación en cascada de un filtro *feedforward comb* con otro *feedback comb* (ver figura 2.16).

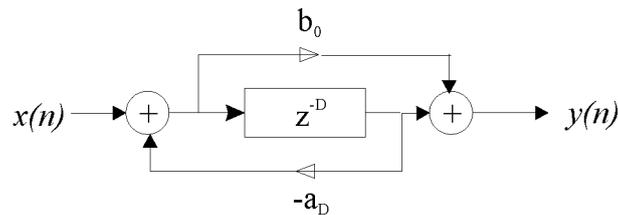


Figura 2.16 Filtro *allpass*.

Su ecuación puede escribirse como:

$$y(n) = b_0 x(n) + x(n-D) - a_D y(n-D) \quad (2.31)$$

Su función de transferencia es:

$$H(z) = \frac{b_0 + z^{-D}}{1 + a_D z^{-D}} \quad (2.32)$$

Y la respuesta en frecuencia:

$$G(\omega) = |H(e^{j\omega})| = 1 \quad (2.33)$$

La característica principal de estos filtros es que, si bien su respuesta en amplitud es unitaria, su respuesta en fase es arbitraria, produciendo una decorrelación de las componentes de la señal.

Diseño de reverberadores

Schroeder (1961) es quien propone, por primera vez, la utilización de líneas de retado en serie para simular los primeros ecos. Y también el uso de filtros *comb* y *allpass* en el diseño de reverberadores, partiendo de dos modelos, destinados a la simulación de la reverberación natural de un ambiente. El primero está basado en cinco filtros *allpass* conectados en serie, donde la señal directa se suma a la salida, escalada por un valor. El segundo, por cuatro filtros *comb* en paralelo, seguidos de dos filtros *allpass* en serie, donde también la señal directa se suma a la salida. La adición de la señal directa, multiplicada por un factor, persigue el control de la relación sonido directo-reverberado, que como sabemos, aporta indicios en la determinación de la distancia fuente-sujeto. Cualquiera sea el modelo, queda establecido que la elección cuidadosa de los valores de ganancia, y tiempos de retardo de los filtros, influye decisivamente en la calidad de los resultados.

Allpass en cascada

En la figura 2.17 se aprecia una sección de tres filtros *allpass* en cascada basada en la propuesta de Schroeder. El valor típico para las ganancias (g) es 0.7, con retardos de 1051, 337 y 113 muestras (números primos).

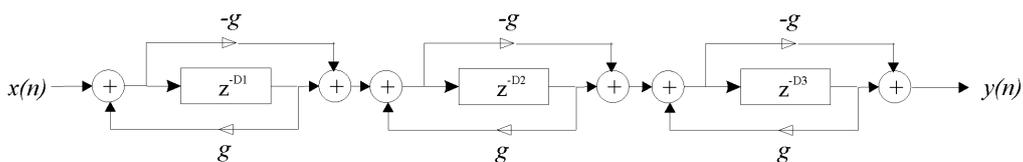


Figura 2.17 Ejemplo de filtros *allpass* en cascada.

***Comb* en paralelo y *allpass* en serie**

Moorer (1979), propone incluir un filtro simple de un polo en la realimentación de los filtros *comb*, con el propósito de atenuar las altas frecuencia en las sucesivas pasadas, e imitar así la acción filtrante del aire y la absorción de las superficies reflectantes. En la figura 2.18 se muestra el diagrama de este filtro *comb* con pasabajos.

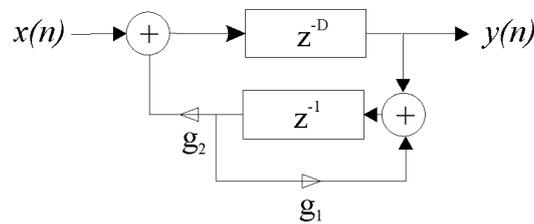


Figura 2.18 Filtro *comb* con pasabajos.

La ganancia general de la realimentación (g) está dada por:

$$g = \frac{g_2}{1 - g_1} \quad (2.34)$$

Si g_1 equivale a 0, entonces el filtro se reduce a un *comb* normal, y mientras más se aproxima a 1, más rápido se atenúan las frecuencias altas respecto a las bajas frecuencias. El factor g , por otra parte, controla el tiempo de reverberación.

La figura 2.19 muestra una combinación de seis filtros *comb* con pasabajos en paralelo y un *allpass* en serie, que surge como modelo luego de extensas experimentaciones.

Los filtros *comb* crean patrones repetitivos que aumentan la densidad de las reflexiones, mientras que el filtro *allpass* provee retardos de fase diferenciados en función de la frecuencia, que afectan a la totalidad de los patrones. La señal directa se mezcla a la salida para controlar la relación sonido directo-reverberado. En los sistemas de reproducción no

monoaurales, suele agregarse un filtro allpass por canal para decorrelacionar las señales, y brindar así una sensación ambiental más realista.

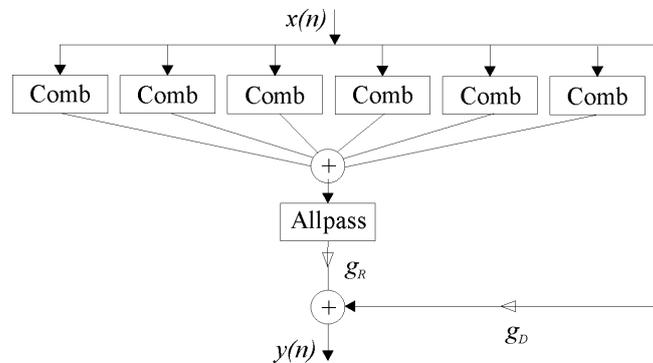


Figura 2.19 Reverberador propuesto por Moorer (1979), con 6 *comb* y 1 *allpass*.

Los valores de ganancia g_1 propuestos por Moorer, para una frecuencia de muestreo de 50 KHz y un tiempo de reverberación de 2 segundos ($g = 0.83$), son 0.46, 0.48, 0.5, 0.52, 0.53 y 0.55. Y los tiempos de retardo de los filtros *comb* son de 50, 56, 61, 68, 72 y 78 milisegundos. La ganancia del *allpass* es 0.7 y el retardo es de 6 ms.

JCRev

John Chowning, partiendo de las premisas planteadas por Schroeder, diseñó un reverberador, denominado *JCRev*, que utiliza tres filtros *allpass* en serie (a diferencia de los cinco propuestos por Schroeder), seguidos por cuatro *feedforward comb* en paralelo. La función de los *allpass* es crear una reverberación densa desprovista de color. Los tiempos de retardo originales de Schroeder, en milisegundos, se obtienen a través de:

$$D^i T = \frac{100}{3^i} \quad \text{para } i = 0, 1, \dots, 4 \quad (2.35)$$

Los valores estimados por Chowning son de 1051, 337 y 113 muestras, con una ganancia de 0.7, según se observa en la figura 2.20. Los filtros *comb* agregan cierto color a la reverberación, y simulan cuatro primeros ecos. Los valores de estos filtros influyen en la percepción de las dimensiones del ambiente.

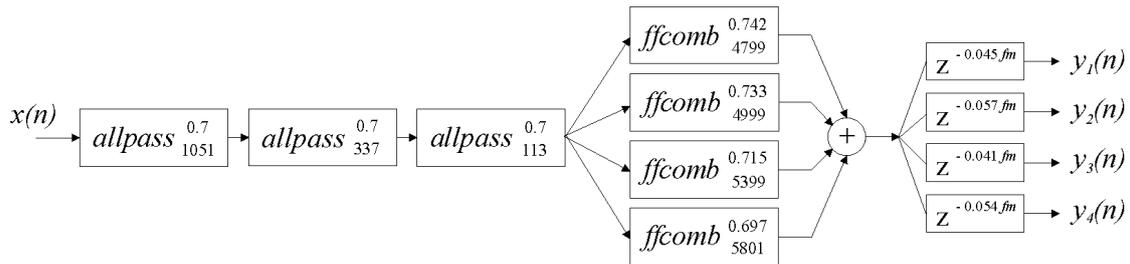


Figura 2.20 Reverberador *JCreV*, de Chowning.

La salida del reverberador es conducida a cuatro canales, decorrelacionados con líneas de retardo. Los tiempos de retardo son relativos a la frecuencia de muestreo, y redondeados al entero más próximo.

FDN (*Feedback delay networks*)

Gerzon (1976) fue el primero en sugerir la aplicación de redes formadas por líneas de retardo realimentadas en el diseño de reverberadores, proponiendo un banco de retardos en paralelo controlados por una matriz de coeficientes de realimentación.

Stautner y Puckette (1982), por su parte, propusieron un modelo similar en cuatro canales, donde una matriz de 4x4 controla el comportamiento resonante del ambiente, interconectando las entradas y salidas de las líneas de retardo. La salida de cada línea se reintroduce en todas las restantes, escaladas por coeficientes de realimentación especificados en la matriz, produciendo una densidad de ecos muy alta.

Jot y Chaigne (1991) generaron una metodología para el diseño de FDN, y estos reverberadores son considerados hoy como el mayor logro en el campo de la reverberación artificial. Sus primeros trabajos se concentraron en un modelo de única entrada y única salida, formado por tres líneas de retardo y un filtro de corrección tonal ($E(z)$) destinado a ecualizar los resultados de aplicar tiempos de caída diferente en función de la frecuencia. El esquema de este reverberador se ilustra en la figura 2.21.

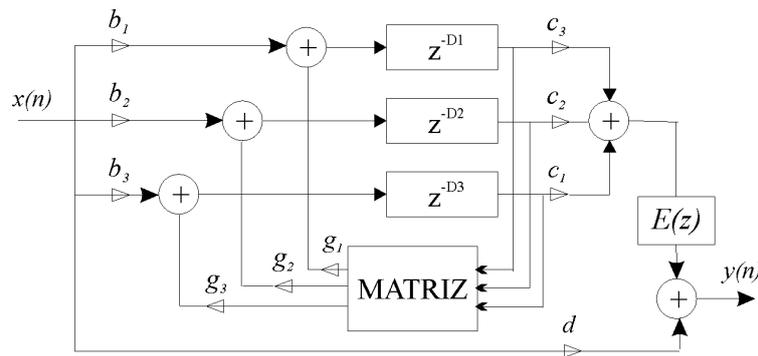


Figura 2.21 Estructura FDN para reverberación propuesta por Jot (1991).

Los tiempos de retardo elegidos son números primos, para minimizar las superposiciones en la respuesta a impulso, y los niveles de realimentación son escogidos de modo tal que aseguren la estabilidad del sistema. En este modelo se independiza el control de la densidad de los ecos y las características espectrales, de la caída de la reverberación.

Un marco equivalente en el desarrollo de reverberadores es la técnica DWN (*Digital Waveguide Network*), definida por un conjunto de líneas de retardo bidireccionales conectadas por nodos, como un modelado físico de tubos interconectados entre sí.

Cualquiera sea el modelo elegido, se considera que la reverberación ideal de un impulso se asemeja a un ruido que decae exponencialmente. Es por esto que en el diseño de reverberadores se comienza por un tiempo de reverberación infinito, que permita modelar la señal de salida convenientemente. El paso siguiente consiste en otorgar el tiempo de reverberación adecuado a cada banda de frecuencia.

Decorrelación

La decorrelación de las señales emitidas por los parlantes de un sistema de reproducción contribuye efectivamente a la creación de imágenes espaciales más realistas, especialmente cuando se trata el problema de la reverberación. Según Kendall (1995), la decorrelación produce, al menos, cinco efectos sobre los eventos auditivos:

- Elimina perceptualmente la coloración producida por la interferencia de las señales afectadas por múltiples retardos
- Produce campos sonoros difusos
- Provoca la externalización de las imágenes percibidas con auriculares
- Relativiza las diferencias debidas a la posición del oyente en el ambiente
- Compensa los efectos negativos del efecto de precedencia.

Hemos visto, al tratar el reverberador de Moorer, que un modo de decorrelacionar las señales consiste en la utilización de filtros *allpass* diferenciados para cada canal. Otro método se basa en la convolución de la señal monofónica con otras dos, que guardan el grado de correlación buscado. La magnitud de la correlación se expresa a través de un coeficiente que varía entre -1 y +1, denominado coeficiente de correlación cruzada, y que se obtiene considerando el valor absoluto más elevado de la función de correlación:

$$\Omega(\Delta t) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} y_1(t) y_2(t + \Delta t) dt \quad (2.36)$$

donde Δt representa a la diferencia temporal que separa a $y_1(t)$ e $y_2(t)$.

Cuando las señales son idénticas el coeficiente equivale a 1, cuando están en oposición de fase equivale a -1, y cuando son diferentes, el valor tiende a cero. Sobre la obtención de las funciones de correlación y su aplicación puede consultarse el artículo de Kendall citado.

Calidad de la reverberación

Se han desarrollado métodos de medición de la calidad de la reverberación basados en el análisis de la respuesta a impulso de un sistema, y algunas de las cualidades a evaluar han sido definidas matemáticamente. Uno de los aspectos a considerar es la “densidad de frecuencia” o “densidad modal”, definida como el número de picos de frecuencia (*eigenfrecuencias*) por Hz. Para un banco de filtros *comb* en paralelo, el índice se obtiene sumando los valores de retardo en segundos, y se considera que un valor de densidad de frecuencia adecuado debe ser mayor a 0.15 (Jot, 1991). En el reverberador de Moorer, este valor se ubica en 0.385, por lo cual supera el mínimo ampliamente. Se considera, además, que la caída de la reverberación debe ser diferenciada por bandas y poseer tiempos distintos en función de la frecuencia, al igual que la absorción debida a la acción atmosférica. La distribución de energía debe ser tal que no produzca efectos de coloración no deseados.

Otro aspecto es la “densidad temporal”, definida por el número de reflexiones por segundo. También es posible calcularla para un banco de *combs* en paralelo, sumando los inversos de los tiempos de retardo. Para el reverberador de Moorer es 95.67, muy alejado de 10000, que es el valor considerado ideal. Pero hay que considerar que este resultado no tiene en cuenta que los filtros *comb* no son los tradicionales, sino que poseen un pasabajos en la realimentación, ni la acción del filtro *allpass*. Las reflexiones deben decrecer exponencialmente, y debe existir un adecuado control del tiempo de reverberación.

La forma en que se produce la caída de la energía en la reverberación es también considerada. Schroeder desarrolló una técnica de medición que permite evaluar la curva de caída de energía, disponiendo las respuestas en frecuencia a lo largo del tiempo. Posteriormente, Jot (1997) modificó esta apreciación a través de una representación tridimensional del fenómeno, denominado EDR (*Energy Decay Relief*).

Parámetros subjetivos de la reverberación

Kendall y Martens (1984) consideran cuatro parámetros subjetivos en la evaluación de la localización en recintos cerrados. El primero se refiere a la “calidad de la distancia”. La percepción de la distancia está basada principalmente en la estimación de la sonoridad en relación al timbre, y en la apreciación de la relación entre sonido directo y reverberado. El modelo de Chowning (1971), por ejemplo, provee una sensación de distancia bastante apropiada, sin embargo, sólo es efectiva para grandes dimensiones del ambiente virtual. En ambientes reducidos, la correcta percepción de este parámetro estaría dada por la estructura de las primeras reflexiones.

La “definición”, por otra parte, se relaciona con la claridad en la percepción de las características de la fuente, como si se tratara del foco de una imagen proyectada a través de una lente. Las imágenes definidas son, a la vez, inteligibles, y tal propiedad depende de las características temporales del sonido, de la reverberación del ambiente, y de la separación de los objetos sonoros en el espacio. La posición de un sonido continuante, percibido en un ambiente reverberante, no se puede determinar con facilidad. Un impulso percibido en una cámara anecoica, por el contrario, es susceptible de una determinación precisa de su posicionamiento.

Otro parámetro subjetivo es la “espaciosidad”, que se presenta como una característica propia del ambiente. A veces es denominada “presencia” o “reverberancia”, y se refiere a la percepción de las dimensiones del espacio y calidad de la reverberación. Un mayor grado de espaciosidad origina una disminución en la definición. Este efecto es el que caracteriza a un buen diseño de una sala de conciertos, donde se persigue una audición pareja desde cualquier ubicación, a expensas de un alto grado de definición espacial de la fuente.

Por último, la “textura espacial” es la interacción del sonido con el ambiente, y está directamente relacionada con la estructura de las reflexiones. Una misma fuente ubicada en dos posiciones, a igual distancia e igual sonoridad, es percibida de dos modos distintos.

Un modelo de reverberador cuyo control está basado en parámetros perceptuales es el diseñado por Jot (1996).

La reverberación incrementa el realismo de los eventos auditivos y ayuda a la externalización en la audición con auriculares, sin embargo, afecta la precisión en la localización. La distorsión producida por los ecos en el espectro percibido depende de la ubicación del oyente en el recinto. Esta distorsión se vuelve significativa cuando el oyente se encuentra cerca de los límites del cuarto, dado que la reflexión sobre la pared más próxima llega inmediatamente después que el sonido directo, produciendo un efecto similar al de un filtro *comb* (Shinn-Cunningham, 2000). Mediciones de las HRTF en ambientes reverberantes, para distintas posiciones del oyente, confirman que el menor nivel de degradación ocurre para las posiciones cercanas al centro del recinto.

2.3.2 Descripción del modelo de Moore

El sistema de espacialización de F. Richard Moore (Moore, 1983) se basa también en la reproducción cuadrafónica del sonido. Los cuatro parlantes se ubican en las esquinas del auditorio, y determinan los límites del “cuarto real”, donde se ubican los oyentes. La posición de la fuente virtual se encuentra más allá de este espacio físico, dentro de lo que denominamos el “cuarto virtual” (ver figura 2.22). Este modelo, que incorpora el tratamiento del ambiente virtual a través de reflexiones tempranas y reverberación, fue ampliamente utilizado en la composición de música electroacústica y mixta.

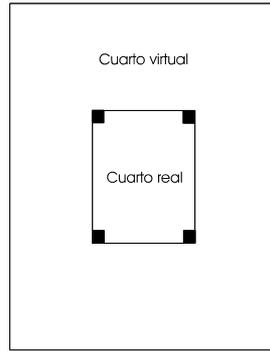


Figura 2.22 Ambiente real y ambiente virtual en el modelo de Moore (1983).

La ubicación de una fuente sonora entre la posición de un parlante y el oyente, en todos los casos que estamos tratando, resulta prácticamente imposible. Para que esto suceda, es preciso simular las diferencias interaurales de intensidad, de tiempo y los indicios espectrales empleando reproducción transaural (que veremos en el capítulo siguiente), acondicionar acústicamente la sala tratando de evitar las reflexiones en su interior, y aún así, la imagen espacial estará sumamente condicionada por la posición del oyente.

El sistema de Moore no es ajeno a la deformación de la imagen en función de la posición del oyente, pero ésta resulta de un modo natural, creando una perspectiva sonora, análoga a la perspectiva visual que cada individuo experimenta dentro de la sala.

El sonido que se produce en forma simulada en el cuarto virtual ingresa al cuarto real a través de cuatro orificios representados por los parlantes. Si imaginamos el movimiento de la fuente sonora en el exterior, obtendremos una idea de su ubicación por simple comparación: la fuente se encuentra cerca del parlante que suena más fuerte. Pero no sólo nos guiamos por la intensidad del sonido, sino también por su tiempo de arribo. Estas diferencias de intensidad y de tiempo entre las señales emitidas por los parlantes es la base de la simulación de este modelo, que posee una fuerte discriminación frente-atrás por tratarse de un sistema cuadrafónico.

No sólo simulamos el sonido directo y las primeras reflexiones que parten de la fuente a cada uno de los orificios (parlantes), con sus niveles de intensidad y tiempos de llegada relativos, sino también la reverberación global, que ayuda a determinar las características de la sala imaginaria. Sólo debemos limitarnos a reproducir, en cada parlante, copias del sonido original (directo y reflejado) con las intensidades y retardos calculados como si la fuente virtual realmente existiera. Vemos en la figura 2.23 una representación –por medio de rayos- del sonido directo y las primeras reflexiones que alcanzan a dos de los parlantes.

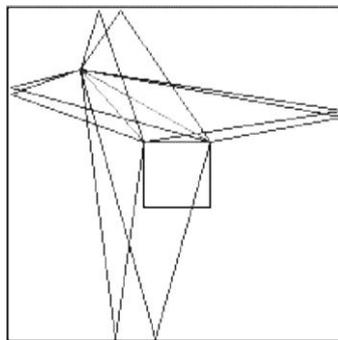


Figura 2.23 Primeras reflexiones que alcanzan la posición de los parlante frontales.

Para obtener las distancias recorridas podemos remitirnos al método de las fuentes imaginarias, tratado anteriormente. La figura 2.24 lo muestra para las cuatro primeras reflexiones que alcanzan la posición del parlante derecho delantero.

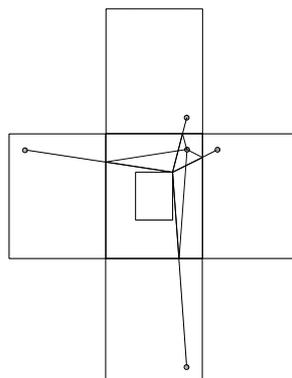


Figura 2.24 Aplicación del método de las fuentes imaginarias en el cálculo de las distancias fuente-parlante, para las señales reflejadas.

A partir de la obtención de la distancia de cada rayo (sonido directo y cuatro primeras reflexiones a cada uno de los cuatro parlantes, dando un total de 20) es posible calcular el retardo y el factor de atenuación de la amplitud.

2.3.3 Implementación en tiempo real

Implementé un modelo simplificado en Max-MSP con propósitos didácticos. En la figura 2.25 se aprecia la sección de cálculo, donde se realizan las operaciones que dan por resultado las distancias, amplitudes y retardos relativos a cada vector en función de la posición de la fuente. Como vimos, se trata de 20 rayos ([sonido directo + 4 primeras reflexiones] * 4 parlantes). Dado que se trata de un número significativo de cálculos, pero todos basados en sólo tres ecuaciones, opté por empaquetar a todas las variables en listas que son leídas por la unidad *vexpr*. Esta unidad lee tantas listas como variables tiene la expresión que ella contiene, y realiza las operaciones tomando los valores secuencialmente, por número de orden dentro de las listas. Los resultados se entregan empaquetados a la salida. Las distancias, por ejemplo, se obtienen por $r = \sqrt{(a-b)^2 + (c-d)^2}$, que tiene como variables a distintas coordenadas (fuentes, fuentes fantasma, posición de los parlantes).

Los retardos son calculados en función de la distancia, en milisegundos con $\tau_x = D_x / .34$, y las amplitudes de cada rayo con la ley del cuadrado inverso. Para corregir perceptualmente la sensación de alejamiento, la potencia a la cual se eleva la distancia se establece como variable (p) en $1 / (1 + r^p)$.

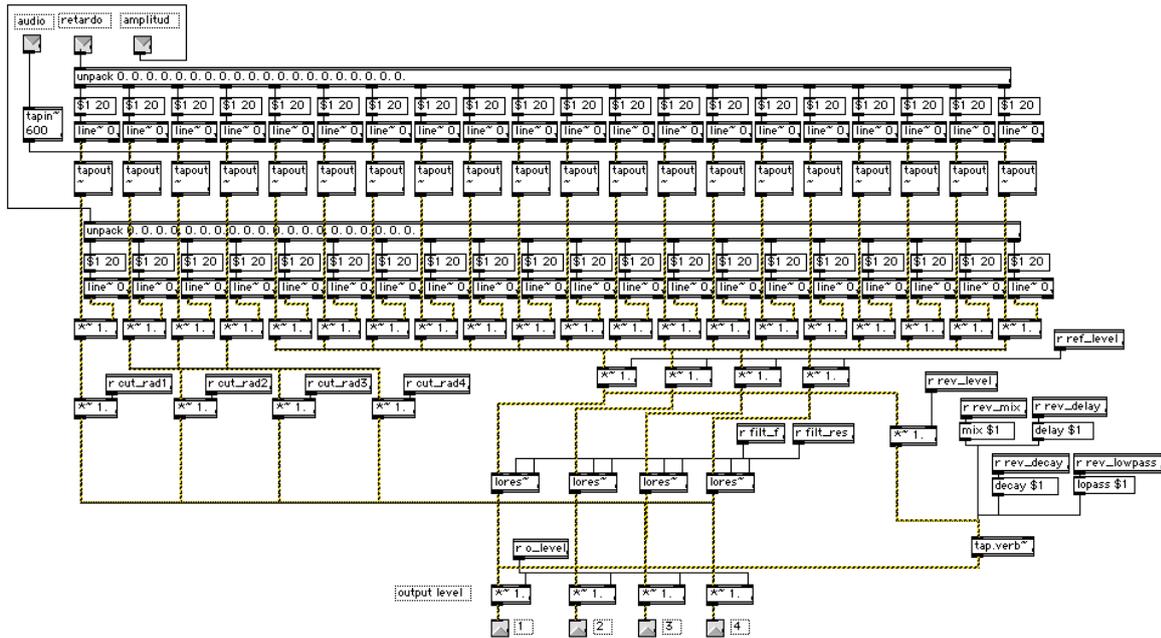


Figura 2.26 Aplicación práctica de los retardos y ganancias en la versión en tiempo real del modelo de Moore (Cetta, 2003a).

El modelo original de Moore incorpora, además, el control del ángulo de difusión de la fuente sonora y el esquema de radiación. La ecuación siguiente genera los factores de amplitud que simulan un diagrama hipercardiode:

$$r(\phi) = \left[1 + \frac{(back - 1)|\theta - \phi|}{\pi} \right]^2 \quad (2.37)$$

donde $r(\phi)$ es un escalar del rayo simulado y ϕ es su dirección, θ es la dirección de la fuente respecto al oyente, y $back$ es el factor de radiación en la dirección contraria a θ .

También considera la eliminación de las reflexiones que atraviesan el cuarto real, aplicándoles un *fade out* de 50 ms. Estas características, por cuestiones de simplificación, no han sido incorporadas en esta versión.

El autor mismo del modelo realizó recientemente una versión en tiempo real (Yadegari, Moore y otros, 2002), optimizando el algoritmo que calcula los retardos, y mejorando el factor de atenuación por difracción cuando las reflexiones intersectan el cuarto real, aplicando:

$$da = \begin{cases} 0 & \text{cuando } TH < ds \\ \left(\frac{TH - ds}{TH} \right)^{CF} & \text{cuando } 0 < ds < TH \end{cases} \quad (2.38)$$

donde da es el factor de atenuación por difracción, ds es la distancia entre la intersección del rayo con la pared y la esquina, TH es el umbral de difracción y CF el exponente del *crossfade*.

2.4 Control vectorial de la amplitud sobre ternas de parlantes

2.4.0 Descripción del modelo de Pulkki

Ville Pulkki (1997) desarrolló una técnica de espacialización basada en paneos de intensidad sobre ternas de parlantes, cuyas ganancias son calculadas de forma vectorial, a la que denominó VBAP (*Vector Base Amplitude Panning*). El campo sonoro es conformado por sectores triangulares delimitados por las ternas (ver figura 2.27) y la diferencia mayor respecto a otras técnicas es que los parlantes pueden estar ubicados en cualquier dirección.

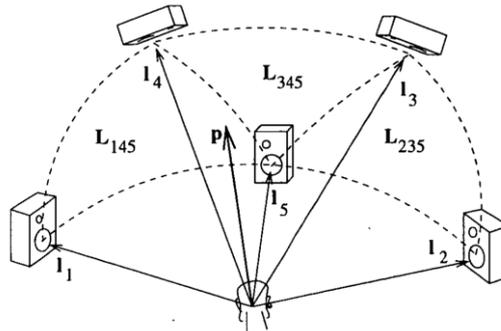


Figura 2.27 Subdivisión de un sector de la esfera virtual que rodea al oyente en ternas de parlantes.

Consideremos en primer término la localización en dos dimensiones. El sistema estereofónico convencional mismo puede ser formulado a través de vectores (ver figura 2.28). La posición de los parlantes queda definida por dos vectores de longitud unitaria $\mathbf{l}_1 = [l_{11} \ l_{12}]^T$ y $\mathbf{l}_2 = [l_{21} \ l_{22}]^T$ (la T denota la transposición de la matriz). El vector $\mathbf{p} = [p_1 \ p_2]^T$ apunta a la fuente virtual, y es una combinación lineal de los vectores a los parlantes:

$$\vec{p} = g_1 \vec{l}_1 + g_2 \vec{l}_2 \quad (2.39)$$

donde g_1 y g_2 son los factores de ganancia de cada canal.

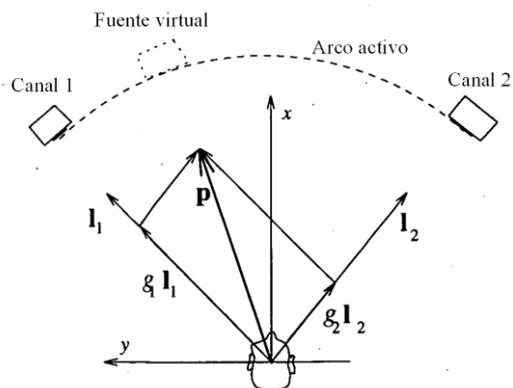


Figura 2.28 El sistema estereofónico definido vectorialmente.

Si escribimos a la ecuación anterior en forma matricial, obtenemos:

$$\vec{p}^T = \vec{g} \vec{L}_{12} \tag{2.40}$$

donde $\mathbf{g} = [g_1 \ g_2]$ y $\mathbf{L}_{12} = [\mathbf{l}_1 \ \mathbf{l}_2]^T$.

$$\vec{g} = \vec{p}^T \vec{L}_{12}^{-1} = [p_1 \ p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1} \tag{2.41}$$

Los valores de g_1 y g_2 obtenidos, satisfacen la ley de la tangente del sistema estereofónico (ecuación 2.4).

Cuando θ_0 es distinto de 45° , las ganancias deben ser normalizadas de acuerdo a:

$$g_{esc} = \frac{\sqrt{k} g}{\sqrt{g_1^2 + g_2^2}} \tag{2.42}$$

para que satisfagan la ecuación (2.0).

Podemos extender este principio a cualquier cantidad de parlantes, distribuidos de forma regular o irregular en el plano horizontal, tomándolos por pares de acuerdo a la posición asignada a la fuente virtual.

De igual forma, si consideramos al sistema en tres dimensiones, son tres los parlantes involucrados (figura 2.29). Los coeficientes de ganancia se resuelven a partir de:

$$\vec{g} = \vec{p}^T \vec{L}_{123}^{-1} = [p_1 \ p_2 \ p_3] \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \quad (2.43)$$

Las componentes del vector \mathbf{g} pueden ser usadas como coeficientes luego del escalamiento, dado por:

$$\mathbf{g}_{esc} = \frac{\sqrt{k} \mathbf{g}}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \quad (2.44)$$

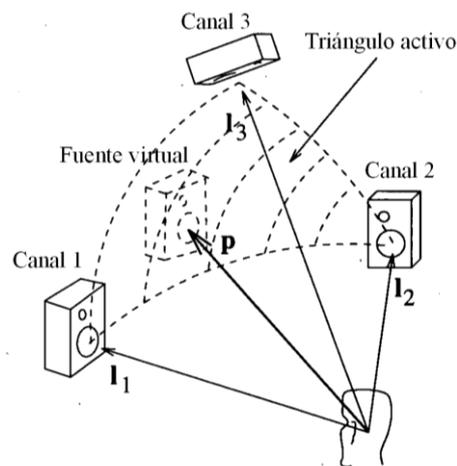


Figura 2.29 Triángulo activo definido por tres canales.

La implementación por software o hardware de esta técnica contempla los siguientes pasos, repetidos cíclicamente:

- 1- Determinar la dirección del vector p .
- 2- Elegir el par o la terna asociada a esa dirección
- 3- Calcular los factores de ganancia
- 4- Realizar un *cross-fade* entre los factores de ganancia anteriores y los nuevos.

En los sistemas de simulación de fuentes virtuales por paneo de intensidad, la fuente fantasma generada por dos, o más canales, posee una imagen extendida, que puede ser erróneamente localizada e incluso puede estar coloreada. Esta situación se vuelve más evidente por comparación, cuando la fuente se desplaza y se detiene en un único parlante, convirtiéndose en una fuente real. Un recurso posible para uniformar la percepción de la imagen espacial es mantener constante el nivel de expansión de la fuente en todas las posiciones, aplicando la misma señal a más de un parlante en todo momento. Pulkki (1999) propuso un método donde la señal es paneada simultáneamente a múltiples direcciones cercanas a la deseada, denominado MDAP (*Multiple Direction Amplitude Panning*). El ángulo mayor de desviación en las direcciones es llamado “ángulo de expansión”. Si empleamos una configuración de parlantes dispuesta sobre el plano horizontal, cuando la dirección de la fuente apunta a un único parlante, el ángulo de expansión está delimitado por dos señales en direcciones próximas que lo rodean (ver figura 2.30).

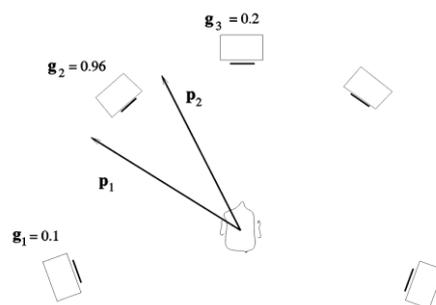


Figura 2.30 Ángulo de expansión en el método MDAP (Pulkki, 1999).

Este método complementa a la técnica VBAP, y persigue compensar la homogeneidad que es posible percibir en otros modelos de espacialización, basados en la reconstrucción del frente de onda (Ambisonics, por ejemplo), donde todos los parlantes emiten señales, independientemente de la posición de la fuente virtual.

2.4.1 Implementación en Max-MSP

Veamos ahora una implementación en Max-MSP, creada por el mismo autor de la técnica (Pulkki, 2000).

El sistema consta de tres objetos interrelacionados: *define_loudspeakers* (define la configuración de parlantes a utilizar), *vbap* (es la unidad de espacialización) y *matrix~* (distribuye las señales a los parlantes). Cada fuente a espacializar precisa un objeto *vbap* propio.

El objeto *define_loudspeakers* se inicializa especificando la posición de los parlantes, y calcula los pares o ternas a partir de esos datos, y la matriz $L_{m,n,k}^{-1}$ para las áreas calculadas. Finalmente, distribuye esta información a todos los objetos *vbap* presentes.

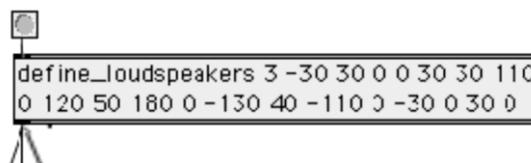


Figura 2.31 Inicialización del objeto *define_loudspeakers* (Pulkki, 2000).

El objeto *vbap* calcula los factores de ganancia dependientes de la posición de la fuente virtual. La información que recibe son los datos de la unidad *define_loudspeakers*, el ángulo de azimut, el ángulo de elevación y el nivel de esparcimiento de la fuente. Al recibir un *bang* ejecuta los cálculos de VBAP y MDAP y envía los factores de ganancia a la salida. Si la

posición de la fuente se define fuera del espacio cubierto por los parlantes, le asigna una posición en el par o terna más cercano.

El valor de esparcimiento de la fuente corresponde al ángulo formado entre la dirección buscada y la dirección agregada más alejada. La cantidad de posiciones calculadas es 7 en 2D y 17 en 3D (ver figura 2.32).

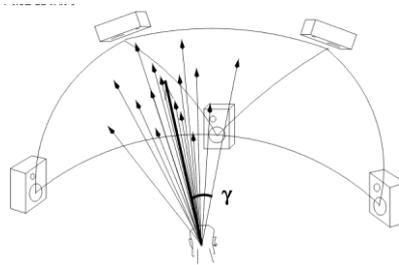


Figura 2.32 Implementación del método MDAP en 3D (Pulkki, 2000).

Los dos objetos mencionados son de control (Max). El único objeto que procesa señales de audio (MSP) es *matrix~*, que recibe las señales de control y las señales a espacializar, y devuelve las señales que alimentan a cada parlante, suavizando las diferencias debidas a los cambios de posición. Entre la salida de *matrix~* y la entrada al conversor digital analógico es posible ubicar un objeto *ls_delays* (figura 2.33), cuya función es retardar las señales destinadas a parlantes que se encuentren más alejados de la esfera base que rodea al oyente.

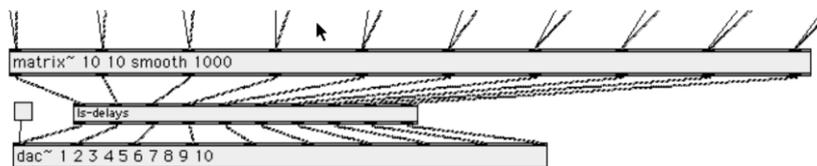


Figura 2.33 Unidad de retardos para compensar las diferencias entre las distancias a los parlantes (Pulkki, 2000).

La simulación corresponde a un ambiente anecoico. Si se desea crear un ambiente virtual, es preciso programar un modelo extendido que considere los primeros ecos y la reverberación.

2.5 Sistemas basados en la reconstrucción del frente de onda

2.5.0 Wave Field Synthesis (WFS)

A fines de los años 80, Berkhout (1988, 1993) propuso una técnica de espacialización del sonido, equivalente acústico de la holografía, que es conocida actualmente como holofonía.

WFS se basa en la reproducción de la curvatura de un frente de onda a partir de matrices de parlantes, y desde el punto de vista físico, parte del Principio de Huygens. Un frente de onda generado por una fuente (la fuente primaria), puede ser simulado a partir de fuentes secundarias distribuidas a lo largo de ese frente, que lo reemplazan (ver figura 2.34)



Figura 2.34. Reconstrucción del frente de onda por medio de *wavelets*.

La expresión matemática de este principio está dada por la integral de Kirchoff-Helmholtz, que puede interpretarse como una distribución continua de fuentes secundarias. Cada fuente está compuesta por otras dos elementales, un dipolo -formado por la presión de la señal- y un monopolo -formado por el gradiente de presión.

En el Principio de Huygens, las fuentes secundarias son consideradas sólo por su amplitud dado que, por tratarse de un frente de onda, las fases son idénticas en toda la superficie. La integral de Kirchoff-Helmholtz, en cambio, contempla cualquier forma geométrica, y la

amplitud y la fase de las fuentes secundarias. Por esto, el principio de Huygens es visto como un caso particular en la integral.

La integral de Kirchoff-Helmholtz expresa que para cualquier punto de escucha ubicado dentro de un volumen V , que no contenga a la fuente, la presión sonora sobre ese punto puede ser calculada si se conoce la presión y su gradiente, sobre la superficie S que envuelve a ese volumen. Este principio puede ser usado para sintetizar un campo sonoro dentro de V , aplicando la presión y su gradiente en forma adecuada sobre la superficie. Sin embargo, para que esto sea realizable es preciso que la superficie pueda degenerar en una línea, y que se pueda discretizar el espacio.

Es posible capturar las características del frente de onda registrándolo con micrófonos dispuestos en una circunferencia alrededor de la posición de la fuente. De acuerdo a la teoría de la holofonía, es necesario utilizar pares coincidentes, formados por un micrófono de presión (monopolo) y otro de gradiente de presión (dipolo). En la reproducción, se disponen pares de parlantes –con características diferentes– distribuidos en idéntica forma que los micrófonos.

Una desventaja resulta de considerar al espacio discretamente, y no como un continuo. Las configuraciones discretas de transductores no pueden registrar ni reproducir correctamente las ondas cuya longitud es inferior a la separación entre ellos, dando lugar a un efecto de *aliasing* espacial.

Lo más destacable de la técnica es que la imagen espacial es percibida para cualquier punto del círculo delimitado por los parlantes. Incluso, el oyente puede desplazarse a través del campo acústico con total libertad, y compartirlo con otros oyentes.

Otra ventaja es que las señales registradas no requieren ningún tipo de tratamiento. Se trata de un proceso físico, producido por interferencia de las ondas generadas por las fuentes secundarias.

Si bien la técnica holofónica clásica resulta poco práctica por la cantidad de canales que requiere, modificaciones posteriores -justificadas con correcciones de la integral de Kirchoff-Helmholtz- simplificaron su implementación. La técnica WFS actual emplea un único transductor en lugar del par antes mencionado, tanto en la etapa de grabación (un solo micrófono, figura en ocho o cardioide) como en la de reproducción (un parlante convencional). Por otra parte, la síntesis puede reemplazar al registro de la fuente primaria, originalmente obtenido con la matriz de micrófonos.

La técnica WFS se encuentra ahora en plena expansión. Uno de los centros de investigación más productivos se encuentra en los laboratorios de France Telecom, donde se realizan tareas de desarrollo, y experimentaciones subjetivas, de las que participan varios socios europeos.

No obstante las mejoras y su eficacia, aún resulta impracticable para una situación de concierto en términos locales, por su elevado costo.

2.5.1 Ambisonics

Este sistema está basado en la descomposición del campo sonoro en armónicos esféricos, que resultan de la traslación de la ecuación de una onda de frente plano a un sistema de coordenadas esféricas. Las primeras cuatro componentes se denominan W, X, Y y Z, y conforman el primer orden del sistema. Las componentes de orden superior -que proveen un grado de precisión mayor en la captación del campo acústico- han sido objeto de estudio en estos últimos años, particularmente en la síntesis de campos cercanos.

Ambisonics (Gerzon, 1973, 1985) es una técnica de grabación y reproducción sonora en tres dimensiones, y a la vez un sistema de espacialización del sonido, muy difundido entre los compositores de música electroacústica. La codificación de las señales en la grabación se realiza utilizando un micrófono especial (*Calrec Soundfield*) formado por cuatro cápsulas direccionales. Las cuatro señales que se registran (Lf, Rb, Rf y Lb) conforman el formato A del sistema.

Conceptualmente, las fuentes virtuales se posicionan en referencia a una esfera de radio unitario, que delimita el campo sonoro. El sistema capta la direccionalidad respecto a un origen. Las fuentes se ubican todas a igual distancia del oyente, como objetos sonoros repartidos sobre la superficie de la esfera.

La codificación, no obstante, puede ser sintetizada partiendo de una señal monofónica, sin necesidad de utilizar micrófono alguno. Las señales bidireccionales se obtienen a partir del producto entre la señal a espacializar y las coordenadas de la fuente. Si especificamos la posición de la fuente en coordenadas esféricas, las ecuaciones son las siguientes:

$$\begin{aligned}
 x(n) &= \cos \theta \cos \varphi s(n) \\
 y(n) &= \sin \theta \cos \varphi s(n) \\
 z(n) &= \sin \varphi s(n) \\
 w(n) &= 0.707s(n)
 \end{aligned}
 \tag{2.45}$$

donde θ es el ángulo de azimut, φ el ángulo de elevación de la fuente y $s(n)$ la señal monofónica a espacializar.

Si expresamos las ecuaciones anteriores directamente en coordenadas cartesianas los cálculos se simplifican, y en consecuencia, el tiempo de procesamiento requerido

$$\begin{aligned}
 x(n) &= x s(n) \\
 y(n) &= y s(n) \\
 z(n) &= z s(n)
 \end{aligned}
 \tag{2.46}$$

Sin embargo, para que la decodificación posterior sea correcta debemos asegurarnos que se cumpla la siguiente regla:

$$x^2 + y^2 + z^2 \leq 1 \tag{2.47}$$

Si la posición de la fuente ocurriera dentro de la esfera de radio unitario, la intensidad general se vería reducida. Para evitar esto, Malham (1998) propone una corrección en la obtención de la señal W:

$$w(n) = \left[1 - 0.293(x^2 + y^2 + z^2) \right] s(n) \tag{2.48}$$

Las cuatro señales obtenidas conforman el denominado formato B de Ambisonic, y se las denomina habitualmente X, Y, Z y W. Las tres primeras albergan las componentes direccionales de los eventos sonoros. La señal omnidireccional W es incluida por varias razones. En principio para compatibilizar el sistema con la reproducción monofónica, y también como medio de extender el área de audición. Sin ella, el espacio útil destinado a los oyentes estaría restringido al centro de la configuración de parlantes.

Puede verse al sistema de codificación de Ambisonics como un tipo particular de paneo de amplitud. Si consideramos el modo pantofónico (en 2D), la señal X contiene a la componente del sonido sobre el eje x , la señal Y a la componente sobre y , y la tercera señal (W) es una mezcla de ambas. Las ganancias que se aplican a las señales son $\cos\theta$, $\sin\theta$ y 0.707, respectivamente. La diferencia con otros sistemas, reside en que las ganancias de las señales X e Y pueden asumir valores negativos, en oposición de fase respecto a W.

Al utilizar un micrófono Ambisonic (ver figura 2.35), las señales obtenidas deben ser transformadas para pasar al formato *B*, de la siguiente forma:

$$\begin{aligned}W &= Lf + Rb + Rf + Lb \\X &= Lf - Rb + Rf - Lb \\Y &= Lf - Rb - Rf + Lb \\Z &= Lf + Rb - Rf - Lb\end{aligned}\tag{2.49}$$

Se desprende de 2.49 que la componente omnidireccional (*W*) está representada por la suma de todas las componentes direccionales.

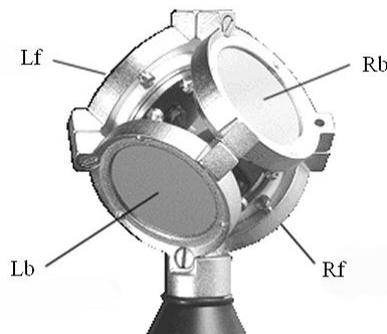


Figura 2.35 Detalle interior del micrófono *Calrec Soundfield*.

Con respecto a la reproducción, distinguimos dos modos:

panofónico: es empleado en la simulación de eventos auditivos sobre el plano horizontal, y requiere un mínimo de cuatro parlantes.

perifónico: considera la elevación de las fuentes virtuales, y requiere un mínimo de ocho parlantes.

El proceso de decodificación persigue la reconstrucción del frente de onda registrado por el micrófono en el momento de grabar, o una versión calculada del mismo, y se realiza para una configuración simétrica de parlantes en particular, previamente elegida. Todos los parlantes se encuentran equidistantes respecto al oyente. La decodificación puede realizarse con distintos niveles de precisión. Para un nivel primario se recurre a las ecuaciones de primer orden, que son las que siguen, para una configuración de parlantes dispuestos en un cuadrado:

$$\begin{aligned}
 s_{fi} &= W + X + Y \\
 s_{fd} &= W + X - Y \\
 s_{ad} &= W - X - Y \\
 s_{ai} &= W - X + Y
 \end{aligned}
 \tag{2.50}$$

Los subíndices identifican al parlante por su ubicación (fi = frente-izquierda, fd = frente-derecha, ad = atrás –derecha y ai = atrás-izquierda). Vemos que la señal Z ha sido descartada, pues la configuración de cuatro parlantes dispuestos horizontalmente no considera la elevación.

Estas ecuaciones de decodificación son adecuadas cuando el área destinada a los oyentes es pequeña en relación al perímetro delimitado por los parlantes. En casos donde esto no se puede garantizar (la posición del oyente es cercana a los parlantes, o está fuera del perímetro) se especifica un *factor de directividad* (para el ejemplo que sigue, igual 0.707) que afecta a las señales X e Y de este modo (Malham y Myatt, 1995):

$$\begin{aligned}
 s_{fi} &= W + 0.707(X + Y) \\
 s_{fd} &= W + 0.707(X - Y) \\
 s_{ad} &= W - 0.707(X - Y) \\
 s_{ai} &= W - 0.707(X + Y)
 \end{aligned}
 \tag{2.51}$$

Al reducir la relación entre W y X, Y y Z mejora la precisión en la localización, pero a la vez se comprime el espacio de audición que resulta convincente. Esta reducción se obtiene cambiando el factor 0.707 por otro mayor, o multiplicando W por una constante menor que 1.

Uno de los principales atractivos de la técnica es la posibilidad de mantener los materiales sonoros codificados, ocupando el espacio necesario para almacenar los cuatro canales (X, Y, Z, W), y decodificarlos para la configuración de parlantes disponible en el lugar de reproducción (4, 6, 8, 10, 12, etc.).

Las señales en formato B pueden ser transformadas de diversas formas. Un proceso típico es la rotación del campo sonoro sobre sus ejes.

Un esquema representativo del sistema puede verse en la figura 2.36.

La codificación de segundo orden (Malham, 1999, 2002) emplea 9 canales (W, X, Y, Z, R, S, T, U y V). Sólo puede obtenerse a partir de la síntesis, dado que hasta el momento no existen micrófonos capaces de obtener ese número de señales. La figura 2.37 muestra los armónicos esféricos, con la denominación en letras mayúsculas recién mencionadas. Las fórmulas correspondientes, en representación azimut-elevación y cartesiana, son las siguientes:

Canal	Esféricas	Cartesianas
W	0.707107	0.707107
X	$\cos \theta \cos \varphi$	x
Y	$\sin \theta \cos \varphi$	y
Z	$\sin \varphi$	z
R	$1.5 \sin \varphi \sin \varphi - 0.5$	$1.5zz - 0.5$
S	$\cos \theta \sin(2\varphi)$	$2zx$
T	$\sin \theta \sin(2\varphi)$	$2yz$
U	$\cos(2\theta) \cos \varphi \cos \varphi$	$xx - yy$
V	$\sin(2\theta) \cos \varphi \cos \varphi$	$2xy$

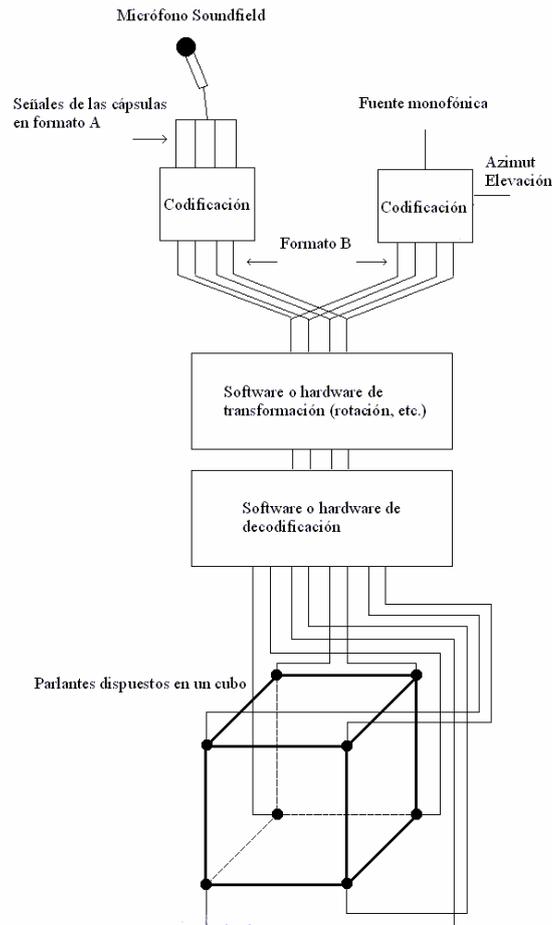


Figura 2.36 Esquema de implementación del sistema Ambisonics.

Veamos ahora matrices de decodificación de primero y segundo orden para configuraciones de parlantes dispuestos en un cuadrado y en un cubo, considerando dos casos. El primero corresponde a la resolución matemática ideal del problema. El segundo, denominado *opuestos controlados*, previene ciertos problemas de fase en las señales de los parlantes y produce, en general, un área de audición mayor a expensas de un menor grado de definición en la localización. Los números, en las tablas siguientes, representan la ganancia de cada señal. Las señales, según vimos antes, se suman, respetando el signo que aparece en cada coeficiente.

Cubo – Primer orden

Parlante	W	X	Y	Z
$S_{fi} \text{ (abajo)}$	0.1768	0.2165	0.2165	-0.2165
S_{fd}	0.1768	0.2165	-0.2165	-0.2165
S_{ad}	0.1768	-0.2165	-0.2165	-0.2165
S_{ai}	0.1768	-0.2165	0.2165	-0.2165
$S_{fi} \text{ (arriba)}$	0.1768	0.2165	0.2165	0.2165
S_{fd}	0.1768	0.2165	-0.2165	0.2165
S_{ad}	0.1768	-0.2165	-0.2165	0.2165
S_{ai}	0.1768	-0.2165	0.2165	0.2165

Cubo – Primer orden – Opuestos controlados

Parlante	W	X	Y	Z
$S_{fi} \text{ (abajo)}$	0.1768	0.0722	0.0722	-0.0722
S_{fd}	0.1768	0.0722	-0.0722	-0.0722
S_{ad}	0.1768	-0.0722	-0.0722	-0.0722
S_{ai}	0.1768	-0.0722	0.0722	-0.0722
$S_{fi} \text{ (arriba)}$	0.1768	0.0722	0.0722	0.0722
S_{fd}	0.1768	0.0722	-0.0722	0.0722
S_{ad}	0.1768	-0.0722	-0.0722	0.0722
S_{ai}	0.1768	-0.0722	0.0722	0.0722

Cubo – Segundo orden

Parlante	W	X	Y	Z	R	S	T	U	V
$S_{fi} \text{ (abajo)}$	0.1768	0.2165	0.2165	-0.2165	0.0000	-0.1875	-0.1875	0.0000	0.1875
S_{fd}	0.1768	0.2165	-0.2165	-0.2165	0.0000	-0.1875	0.1875	0.0000	-0.1875
S_{ad}	0.1768	-0.2165	-0.2165	-0.2165	0.0000	0.1875	0.1875	0.0000	0.1875
S_{ai}	0.1768	-0.2165	0.2165	-0.2165	0.0000	0.1875	-0.1875	0.0000	-0.1875
$S_{fi} \text{ (arriba)}$	0.1768	0.2165	0.2165	0.2165	0.0000	0.1875	0.1875	0.0000	0.1875
S_{fd}	0.1768	0.2165	-0.2165	0.2165	0.0000	0.1875	-0.1875	0.0000	-0.1875
S_{ad}	0.1768	-0.2165	-0.2165	0.2165	0.0000	-0.1875	-0.1875	0.0000	0.1875
S_{ai}	0.1768	-0.2165	0.2165	0.2165	0.0000	-0.1875	0.1875	0.0000	-0.1875

Cubo – Segundo orden – Opuestos controlados

Parlante	W	X	Y	Z	R	S	T	U	V
S_{fi} (abajo)	0.1768	0.1140	0.1140	-0.1140	0.0000	-0.0369	-0.0369	0.0000	0.0369
S_{fd}	0.1768	0.1140	-0.1140	-0.1140	0.0000	-0.0369	0.0369	0.0000	-0.0369
S_{ad}	0.1768	-0.1140	-0.1140	-0.1140	0.0000	0.0369	0.0369	0.0000	0.0369
S_{ai}	0.1768	-0.1140	0.1140	-0.1140	0.0000	0.0369	-0.0369	0.0000	-0.0369
S_{fi} (arriba)	0.1768	0.1140	0.1140	0.1140	0.0000	0.0369	0.0369	0.0000	0.0369
S_{fd}	0.1768	0.1140	-0.1140	0.1140	0.0000	0.0369	-0.0369	0.0000	-0.0369
S_{ad}	0.1768	-0.1140	-0.1140	0.1140	0.0000	-0.0369	-0.0369	0.0000	0.0369
S_{ai}	0.1768	-0.1140	0.1140	0.1140	0.0000	-0.0369	0.0369	0.0000	-0.0369

Las que siguen, son las ecuaciones de codificación de tercer orden (Malham, 2002)

Canal	Esféricas	Cartesianas
W	0.707107	0.707107
X	$\cos(A)\cos(E)$	x
Y	$\sin(A)\cos(E)$	y
Z	$\sin(E)$	z
R	$1.5\sin^2(E)-0.5$	$1.5zz-0.5$
S	$\cos(A)\sin(2E)$	2zx
T	$\sin(A)\sin(2E)$	2yz
U	$\cos(2A)\cos^2(E)$	xx-yy
V	$\sin(2A)\cos^2(E)$	2xy
K	$\sin(E)(5\sin^2(E) - 3)/2$	$Z(5Z^2-3)/2$
L	$8\sin(A)\cos(E)(5\sin^2(E) - 1)/11$	$8Y(5Z^2-1)/11$
M	$8\cos(A)\cos(E)(5\sin^2(E) - 1)/11$	$8X(5Z^2-1)/11$
N	$\sin(2A)\sin(E)\cos^2(E)$	XYZ
O	$\cos(2A)\sin(E)\cos^2(E)$	$Z(X^2-Y^2)/2$
P	$\sin(3A)\cos^3(E)$	$Y(3X^2-Y^2)$
Q	$\cos(3A)\cos^3(E)$	$X(X^2-3Y^2)$

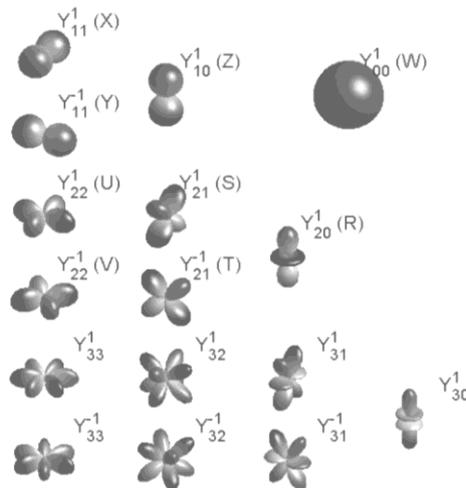


Figura 2.37 Representación de los armónicos esféricos.

Di Liscia (1997, 1999) implementó la técnica Ambisonics de codificación de primer orden y decodificación cuadrifónica, en un software para sistema operativo Windows, que denominó WDSPA. El programa está basado en la simulación del efecto Doppler, la generación de las primeras reflexiones en relación a las dimensiones del recinto y la aplicación de reverberación global con control de las características ambientales. La definición de la posición de la fuente y las trayectorias espaciales se realiza a través de una interfaz gráfica de múltiples posibilidades, que incluye escalamiento traslación y rotación de los recorridos especificados, y el control de la velocidad del movimiento. La figura 2.38 muestra una trayectoria diseñada, haciendo uso de la interfaz mencionada.

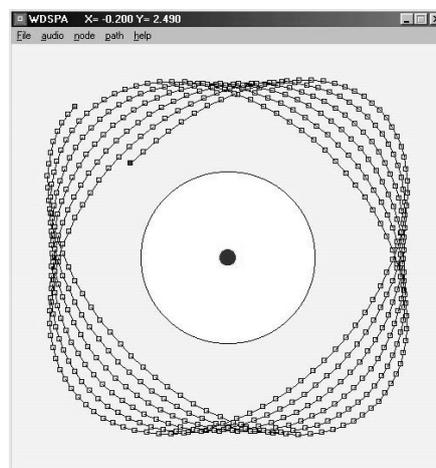


Figura 2.38 Interfaz gráfica par el trazado de trayectorias en el programa WDSPA (Di Liscia, 1999).

2.5.2 Implementación en tiempo real

Programé varios objetos en lenguaje “C” para espacialización del sonido con Ambisonics, en el entorno Max-MSP. La figura 2.39 muestra la pantalla de ayuda del codificador de primer orden, que devuelve las señales W, X, Y y Z, y acepta en sus *inlets* la señal monofónica a espacializar, y las coordenadas cartesianas tridimensionales de la posición de la fuente. Las señales de salida están escaladas por la distancia. En el ejemplo, se descarta la señal Z, dado que se decodifica para un arreglo de cuatro parlantes sobre el plano horizontal, donde no es posible simular la elevación.

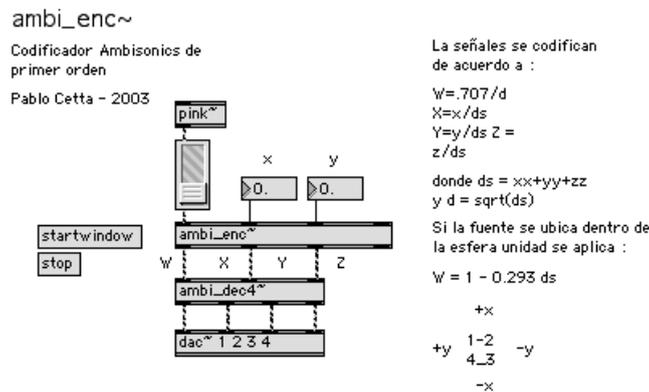


Figura 2.39 Codificador Ambisonics de primer orden en Max-MSP.

La figura 2.40 está referida al decodificador mencionado.

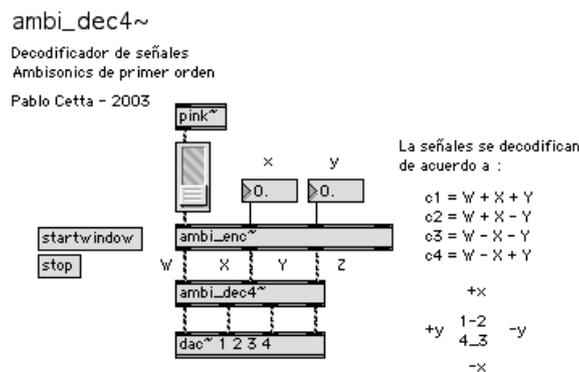


Figura 2.40 Decodificador Ambisonics de primer orden, para cuatro parlantes dispuestos horizontalmente.

Utilizando el mismo codificador, es posible decodificar para diferentes configuraciones de parlantes. En la figura 2.41 se muestra el que corresponde a ocho parlantes dispuestos en un cubo.

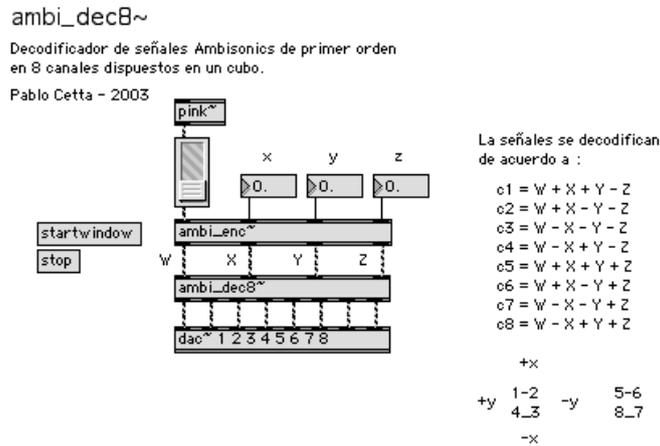


Figura 2.41 Decodificador Ambisonics de primer orden, para ocho parlantes dispuestos en un cubo.

La figura siguiente muestra al codificador de segundo orden. Cuando se decodifica para cuatro parlantes, como en el ejemplo, la única señal que se agrega es la V.

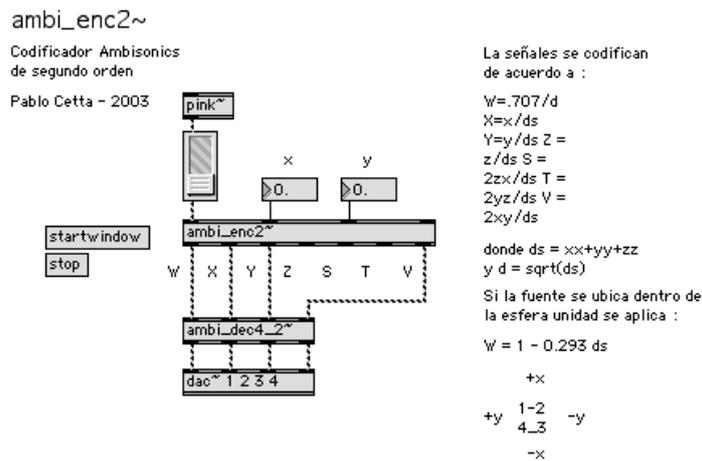


Figura 2.42 Codificador Ambisonics de segundo orden.

La figura 2.43 corresponde al decodificador utilizado en el ejemplo anterior.

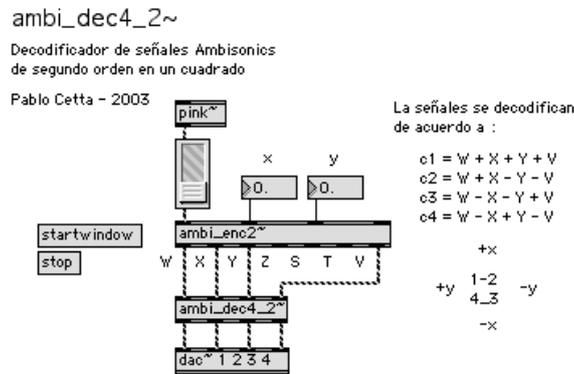


Figura 2.43 Decodificador Ambisonics de segundo orden para cuatro parlantes dispuestos horizontalmente.

Y por último, el decodificador de segundo orden para ocho parlantes dispuestos en un cubo. Las señales utilizadas son W, X, Y, Z, S, T y V.

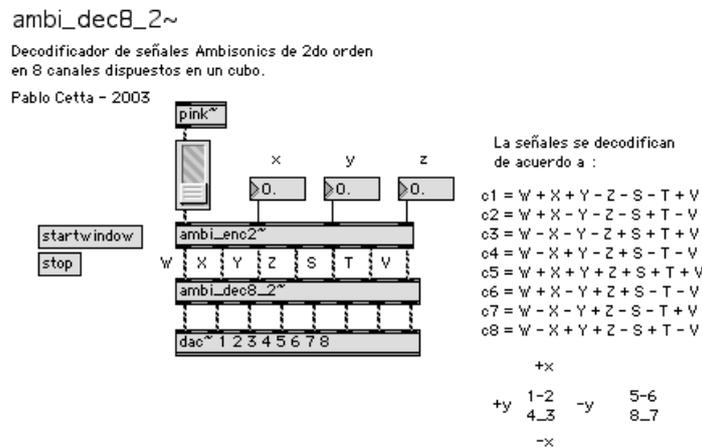


Figura 2.44 Decodificador Ambisonics de segundo orden para ocho parlantes dispuestos en un cubo.

Antes de finalizar con este capítulo vamos a comentar un desarrollo importante en el área de la localización espacial del sonido, que incluye entre otras, a la técnica Ambisonics: el espacializador del IRCAM. El proyecto *Spatialisateur* (Jot, 1996) surge de las investigaciones sobre las caracterizaciones objetivas y perceptivas de las cualidades acústicas de las salas realizadas en el IRCAM (Julien y otros, 1992), y los aportes sobre reproducción tridimensional de ambientes sonoros desarrollados por Espaces Nouveaux, que incorporan

tecnología desarrollada en Telecom de París. El software, concebido como una estructura modular, fue implementado en Max-MSP sobre un objeto principal denominado *spat* ~.

Los cuatro módulos que conforman el sistema son:

source~: es el encargado del tratamiento de las señales de entrada, e incluye ecualización, simulación del efecto Doppler con una línea de retardo, y simulación de la acción absorbente del aire con un filtro pasabajos.

room~: es un reverberador multicanal basado en FDNs, que simula los efectos que ocurren en el ambiente. Emplea algoritmos realizados por Jot para France Telecom, empresa propietaria de las licencias.

pan ~: es el encargado de procesar las señales de acuerdo a la técnica de localización elegida, y controlar la posición de las fuentes virtuales. Las técnicas de localización son estéreo convencional, estéreo 3/2 derivada de los sistemas usados en cine (Thiele, 1993), cuadrafonía u octofonía (con Ambisonics, por ejemplo) y reproducción binaural o transaural

out ~: funciona como decodificador de la técnica de espacialización elegida en el módulo *pan* ~ y realiza una ecualización final de las señales destinadas a los parlantes.

Cuando se diseña la simulación de ambientes virtuales, la estrategia de control de las cualidades acústicas se orienta, en general, al control de la geometría del lugar, considerando particularmente los límites del recinto y los obstáculos a la libre propagación en su interior, teniendo en cuenta, además, la posición relativa del oyente y de la fuente, y la orientación y directividad de la emisión sonora. En este espacializador, en cambio, el control acústico de la sala se realiza a través de factores perceptivos mutuamente dependientes, que reemplazan a las variables físicas. De este modo, el factor denominado “reverberación tardía” afecta al

tiempo de reverberación, la “vivacidad” de la sala afecta a las variaciones temporales de la reverberación con la frecuencia, la “presencia” de la fuente se relaciona con la energía del sonido directo y las primeras reflexiones, y la “brillantez” con la energía de los primeros ecos según la frecuencia.

Hemos visto hasta ahora las principales técnicas de espacialización del sonido y su aplicación práctica, a partir de la programación de computadoras. Omitimos referirnos al sistema binaural, ya que será tratado en el capítulo siguiente, como parte de los principios en los cuales se inspira el modelo propuesto para esta tesis.

2.6 Conclusiones

Del análisis de las técnicas de espacialización, y de la experimentación con los modelos propuestos, adaptados al ámbito del procesamiento de sonido y música en tiempo real, surgen algunas observaciones a considerar.

En todos los casos, el control de áreas de audición extensas, propias de una situación de concierto, se convierte en un problema. Los oyentes alejados del centro, o próximos a alguno de los parlantes, perciben imágenes espaciales severamente distorsionadas. Por otra parte, los eventos auditivos ocurren en un espacio virtual alejado del oyente, que se extiende más allá del perímetro formado por los parlantes. Esta imposición, dada por la imposibilidad de generación de eventos auditivos dentro del espacio de audición condiciona a las aplicaciones musicales.

Si el área de audición es grande, las fuentes aparentes se ubican en posiciones que perceptualmente resultan demasiado alejadas, particularmente cuando se busca resaltar detalles de las estructuras sonoras, o crear situaciones de cierta interioridad. La técnica

Ambisonics, por estar basada en la reconstrucción del frente de onda, mejora este aspecto, pero sólo para los oyentes que se encuentran ubicados en el área central de reproducción.

En el modelo de Chowning, por ejemplo, la sensación de distancia es bastante efectiva cuando las dimensiones del ambiente simulado son muy extensas, caso contrario, induce a cierta indeterminación en la localización de este parámetro. En su momento, la simulación del efecto Doppler le otorgó cierto realismo, pero esta característica puede incorporarse hoy a cualquier sistema.

El modelo de Moore presenta un nivel de efectividad mayor respecto al modelo anterior, dado en principio por la incorporación de las primeras reflexiones a la simulación. No obstante, cuando la fuente se ubica en una posición cercana a alguno de los parlantes, la sonoridad se incrementa desproporcionadamente. Con pruebas informales, es posible comprobar que la sensación espacial se torna más efectiva cuando el oyente es advertido acerca de los principios que rigen al sistema, especialmente cuando considera a los parlantes como si se tratara de orificios en las paredes que separan al ambiente real del ambiente virtual. Esa imagen visual se corresponde de manera efectiva con lo que se escucha.

El sistema Ambisonics, decodificado para cuatro parlantes dispuestos en un cuadrado, presenta un nivel de definición mayor respecto a un sistema cuadrafónico convencional. No obstante, también resulta incapaz de reproducir escenas sonoras tridimensionales. La codificación y decodificación de señales se realiza a través de cálculos muy simples, que requieren una capacidad de procesamiento ínfima. Esto lo convierte en un sistema útil para aplicaciones musicales en tiempo real. Resulta ventajoso, además, mantener codificadas las señales que conforman una obra musical, y decodificarlas para diferentes matrices de parlantes, según las condiciones de reproducción que se presentan en cada situación de concierto.

El método VBAP, presenta una gran robustez en su aplicación. Al igual que Ambisonics, demanda poca capacidad de procesamiento, y permite la utilización de distintas configuraciones de parlantes en la reproducción sonora. Incluso, presenta una ventaja respecto a Ambisonics, ya que los parlantes pueden ubicarse en posiciones arbitrarias. También es comparable en cuanto a la cantidad de canales de audio requeridos para una representación tridimensional del sonido. No obstante, cuatro canales resultan insuficientes para una utilización efectiva, incluso sobre el plano horizontal, si se pretende cubrir los 360°.

Tanto VBAP como Ambisonics resultan eficientes para la mayoría de las aplicaciones musicales, particularmente cuando incorporan la simulación de las características propias de la audición en recintos –absorción del aire, primeras reflexiones y reverberación espacial. Pero un problema adicional se presenta cuando la obra debe ser editada en disco compacto, dado que este soporte responde al sistema estereofónico, de dos canales. Las técnicas binaurales, que veremos en el siguiente capítulo, ayudan a resolver parte de este conflicto a través de la simulación de parlantes virtuales por síntesis binaural. El resultado de este proceso puede ser reproducido en un sistema estéreo convencional.

Capítulo 3

Diseño de programas asistentes para el tratamiento del espacio y composición musical

Capítulo 3

Diseño de programas asistentes para el tratamiento del espacio y composición musical

3.0 Introducción

En este capítulo vamos a describir la técnica binaural y explicar el desarrollo del modelo de localización espacial del sonido propuesto para esta tesis, basado en estos principios. Se trata de un espacializador cuadrafónico de síntesis binaural, con reproducción doble transaural.

También trataremos el diseño de herramientas de procesamiento en tiempo real, destinadas al control del espacio interno y externo de la música. Estos programas -utilizados en la creación de la obra *Interiores*, para cinco instrumentos con procesamiento del sonido en tiempo real- surgen de la necesidad de vincular los parámetros musicales con los recorridos de las fuentes aparentes en el espacio. Además, con efectos de compresión y dilatación del tiempo virtual de la música y del espacio imaginario en el que ésta ocurre.

Finalmente, veremos algunos aspectos formales de *Interiores*, con el propósito de identificar en la obra la aplicación de algunos principios de integración entre Música y Espacio.

3.1 Técnicas binaurales

3.1.0 Principios de grabación binaural

Las grabaciones binaurales se realizan utilizando una cabeza artificial, que captura las señales que ingresan directamente a cada uno de los oídos, con dos micrófonos ubicados en la entrada de los canales auditivos. De este modo, se registran las señales provenientes de las fuentes, también la información espacial vinculada con la posición de los eventos sonoros, y las características del ambiente relativas a la ubicación del maniquí. Se graban las diferencias interaurales de intensidad, de tiempo, y los indicios espectrales, como así también las reflexiones y difracciones que tienen lugar en los pabellones auditivos, la cabeza y el torso. Según vimos, esta información es la que utiliza el sistema perceptual para crear la imagen auditiva.

En la reproducción con auriculares, las señales registradas pueden ser restituidas en los canales auditivos, o fuera de ellos. En ambos casos, es necesaria una etapa de ecualización para compensar las distorsiones producidas por los transductores de la cadena electroacústica, principalmente micrófonos y auriculares. Cuando la reproducción se realiza fuera de los canales auditivos, debemos considerar una ecualización adicional, destinada a contrarrestar la segunda acción de los pabellones durante el proceso de audición (ver figura 3.1).

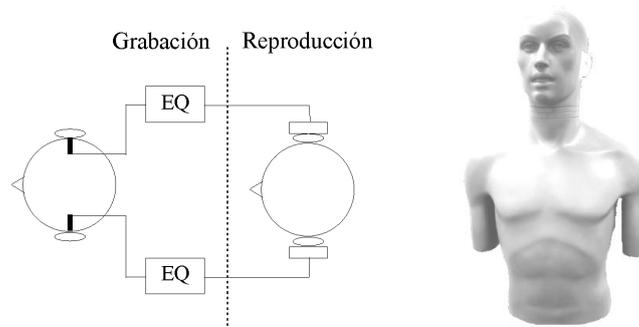


Figura 3.1 Grabación binaural utilizando una cabeza artificial y ecualizaciones asociadas.

3.1.1 Síntesis binaural

La síntesis binaural consiste en el procesamiento de una señal sonora monofónica, a la que se le atribuyen artificialmente diferencias interaurales particulares, como si hubiera sido grabada con una cabeza artificial. El resultado de la síntesis binaural es similar al producido al grabar un sonido cualquiera con una cabeza artificial, ubicando un parlante como fuente sonora en un lugar determinado del ambiente. Se trata de una simulación basada en el proceso de convolución, entre la señal a espacializar y las HRTF correspondientes a la dirección buscada.

3.1.2 HRTF medidas

Es posible catalogar a las técnicas de espacialización basadas en principios binaurales de acuerdo al modelo de HRTF que implementan.

Por una parte, ubicamos las técnicas que utilizan HRTF medidas, ya sea sobre un sujeto en particular, una cabeza artificial, o un grupo de sujetos cuyo valor promedio es calculado.

Las HRTF se obtienen a partir de la medición de respuestas a impulso en ambos oídos, para distintas posiciones de la fuente, a las que posteriormente se les aplica la transformada de Fourier.

Dentro de esta categoría, distinguimos:

- a) HRTF estándar: representan al promedio de una población extensa. No existen aún HRTF normalizadas, sino que cada desarrollo fija su propia norma.

- b) conjuntos de HRTF: formados por grupos reducidos que representan a diferentes tipologías, de los cuales se toman las que mejor representan a un individuo en particular.
- c) HRTF individualizadas: son medidas sobre un único sujeto, a quien va destinado el proceso de localización espacial del sonido.

A través de las técnicas binaurales es posible obtener imágenes en tres dimensiones de gran precisión y realismo utilizando procedimientos relativamente simples. Sin embargo, dada la diversidad morfológica de los individuos, las HRTF medidas sobre una cabeza que no es la propia pueden resultar confusas para algunos oyentes. Las dimensiones, orientación y forma de los pabellones auditivos, el tamaño de la cabeza, el grado de simetría, son todos factores que marcan físicamente las diferencias. En general, las personas localizan mejor la fuente con sus propias HRTF que con las de otras, sin embargo, la percepción puede mejorar cuando se toman como modelo las HRTF de un individuo calificado como un buen localizador. Esto último ocurre en los casos donde el tamaño de la cabeza es muy similar.

Una deficiencia binaural típica se manifiesta en la percepción de imágenes intercraneales, debidas a espectros que no resultan familiares al oyente, o a la ausencia de indicios relacionados con la acústica del ambiente.

Los errores de localización más habituales ocurren para las posiciones frontales, a veces percibidas en la zona posterior o encima de la cabeza. En el diseño de aplicaciones experimentales donde la precisión en la localización de la fuente se vuelve un factor esencial, las HRTF son medidas directamente sobre el destinatario de esa aplicación, para asegurar una correcta generación de imágenes espaciales.

El IMSC (*Integrated Media Systems Center, USC*), por ejemplo, desarrolla un método de clasificación de imágenes de pabellones auriculares almacenadas en una base de datos junto a las HRTFs asociadas. Un sistema de reconocimiento visual, basado en una modificación del algoritmo que se emplea para reconocimiento de la fisonomía de un sujeto, puede buscar las funciones de transferencia que mejor se adapten a un individuo en particular (Kyriakakis, 1998).

En el campo de la realidad virtual la situación presenta un mayor grado de complejidad. Si el oyente mueve su cabeza, en la audición con auriculares, la imagen espacial se desplaza junto a él. Suponiendo que existe una fuente virtual justo enfrente del oyente, al rotar su cabeza, la fuente se desplaza acompañando el movimiento debido a que las diferencias interaurales permanecen constantes. Para evitar este problema se torna necesario detectar los movimientos de la cabeza y actuar en correspondencia, rotando las imágenes espaciales para compensar este efecto. Los sistemas de seguimiento (*head-tracking*) conforman un área particular dentro del campo de la localización con técnicas binaurales, y están basados en principios magnéticos, de ultrasonido, rayos infrarrojos, lasers, o reconocimiento de imágenes faciales en tiempo real. Es obvio que la implementación de tales dispositivos resulta impracticable en una situación de concierto.

3.1.3 HRTF modeladas

Siguiendo con nuestra clasificación, las HRTF modeladas son generadas a través de modelos físicos cuyos parámetros pueden ser ajustados para cada individuo en particular. Si bien existen varias formas de alcanzar el mismo objetivo, vamos a mencionar brevemente a una sola de estas técnicas, conocida como *modelos estructurales*.

El modelado físico de cada una de las partes que componen la antena acústica formada por los pabellones auditivos, la cabeza, el torso y los hombros, se integran conformando funciones de transferencia con parámetros ajustables (ver figura 3.2).

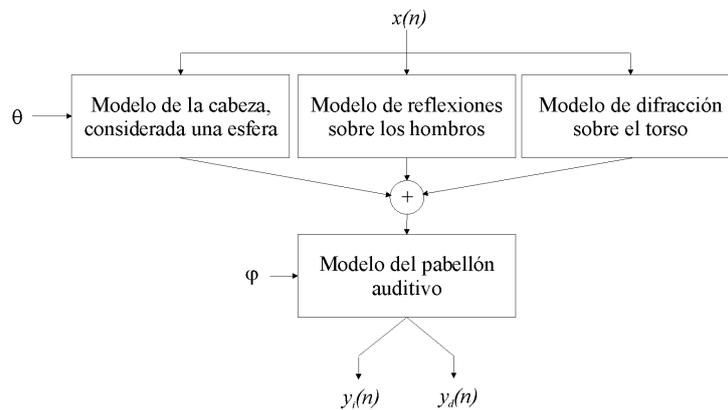


Figura 3.2 Componentes que integran el modelado físico de las HRTF

Un procedimiento simple y efectivo de modelado de las HRTF consiste en considerar a la cabeza como una esfera rígida, según ya vimos en el primer capítulo. La simulación de la diferencia interaural de tiempo es implementada a través de un filtro FIR, que introduce un retardo diferenciado en cada oído, cuya función se expresa a través de $T_d(\theta + \pi/2)$ para el oído izquierdo y $T_d(\theta - \pi/2)$ para el derecho, con

$$T_d = \begin{cases} \frac{a}{c}(1 - \cos \theta) & \text{si } |\theta| < \frac{\pi}{2} \\ \frac{a}{c}(|\theta| + 1 - \frac{\pi}{2}) & \text{si } \frac{\pi}{2} < |\theta| < \pi \end{cases} \quad (3.0)$$

donde a es el radio de la cabeza y c la velocidad del sonido.

También es posible simular la diferencia interaural de intensidad partiendo de un filtro IIR (*infinite impulse response*). Su función de transferencia enfatiza las altas frecuencia cuando la fuente se ubica a 0° y las elimina para un azimut de 180° , simulando así la acción separadora de la cabeza:

$$H(r, \theta) = \frac{\alpha(\theta)r + \beta}{r + \beta} \quad (3.1)$$

donde $\alpha(\theta) = 1 + \cos \theta$ y $\beta = 2c/a$

Adaptando al ángulo de azimut a la posición de ambos oídos, las funciones de transferencia del filtro destinado al oído izquierdo y derecho, respectivamente, se definen por $H_i(r, \theta + \pi/2)$ y $H_d(r, \theta - \pi/2)$.

La simulación de la acción del pabellón auditivo no resulta tan simple, pero es posible lograrla mediante filtros adecuados.

El modelado físico combinado de las difracciones sobre el torso, las reflexiones sobre los hombros, la resonancia del conducto auditivo, e incluso del ambiente, conducen a la optimización del sistema. Un modelo geométrico interesante de la cabeza y el torso puede consultarse en Algazi y otros (2002).

3.1.4 Aplicación de la convolución con HRTF medidas

La convolución es un proceso matemático, utilizado para definir la transformación de una señal que ingresa a un sistema lineal, invariante en el tiempo. En audio digital, este proceso es frecuentemente empleado para simular el efecto de la reverberación. La salida del reverberador es el resultado de la convolución entre la señal a reverberar y la respuesta a impulso del ambiente a simular. El efecto es sumamente preciso, dado que la convolución

aplica el comportamiento de la respuesta a impulso a cada una de las muestras de la señal de entrada.

Existen diversos métodos para la implementación práctica de la convolución. El método lineal es un proceso de filtrado, descrito a través de la ecuación 1.4. En la simulación de la reverberación, $x(n)$ representa a la señal de entrada, $h(n)$ a la respuesta a impulso de la sala, e $y(n)$ a la señal reverberada. La implementación de la convolución lineal se realiza por medio de un filtro FIR (*finite impulse response*) como el que se muestra en la figura 3.3.

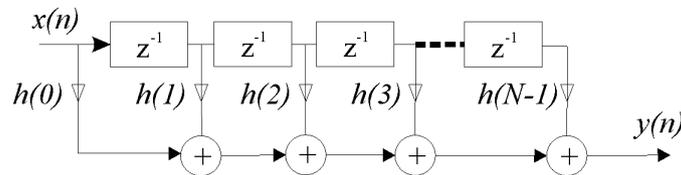


Figura 3.3 Implementación de filtro FIR para convolución lineal.

Si la longitud en muestras de $x(n)$ es n_x , y la longitud de $h(n)$ es n_h , la cantidad de muestras de la señal de salida es $n_x + n_h - 1$. Este incremento de longitud en la salida, respecto a la entrada, surge de la respuesta del sistema a la última muestra ingresada. Si el proceso emula a un reverberador, la diferencia en muestras sería comparable a la caída de la reverberación una vez que el estímulo cesó.

El resultado se produce en tiempo real, pero con un costo elevado de procesamiento, dado que la operación requiere una cantidad de multiplicaciones y sumas igual al producto de la cantidad de muestras de la señal de entrada y la respuesta a impulso ($n_x \cdot n_h$).

Si consideramos que la convolución en el dominio del tiempo es equivalente a la multiplicación en el dominio de la frecuencia, podemos arribar por otra vía al mismo

resultado. La operación ahora consiste en realizar la transformada discreta de Fourier (DFT) de la señal de entrada y la DFT de la respuesta a impulso, multiplicar a ambas, y obtener la DFT inversa para regresar al dominio del tiempo. Para que la cantidad de muestras del resultado sea la misma que en el método anterior, debemos igualar la longitud de ambas señales, agregando $n_h - 1$ ceros a la señal de entrada y $n_x - 1$ ceros a la respuesta a impulso. Si la cantidad de muestras coincide con una potencia de 2, es posible aplicar el algoritmo de la transformada rápida de Fourier (FFT), y si esto no sucede, debemos agregar más muestras en cero a ambas señales hasta alcanzar la potencia de 2 más cercana.

$$y(n) = FFT^{-1}(FFT(x).FFT(h)) \quad (3.2)$$

Si bien el costo de procesamiento se reduce considerablemente, la implementación de este método no es posible en tiempo real, dado que debemos conocer de antemano la señal de entrada en su totalidad.

Otro método, similar al anterior, no considera a la señal de entrada de principio a fin, sino por bloques. La salida se obtiene a través de la superposición y la suma de los resultados parciales. Para comprender el proceso vamos a valernos de un ejemplo.

Suponiendo que la respuesta a impulso consta de 128 muestras, si convolucionamos linealmente igual cantidad de la señal de entrada con la respuesta a impulso, deberíamos obtener un resultado de 255 muestras ($n_x + n_h - 1$). Para que el método de la multiplicación de los espectros de un resultado correcto debemos agregar ceros a ambas señales hasta alcanzar ese número, realizar las FFT, multiplicar los números complejos de ambos resultados, y por último, aplicar la transformada inversa. Las primeras 128 muestras del resultado van directamente a la salida, mientras que las restantes se almacenan para sumarse luego a las primeras 128 muestras resultantes del segundo bloque, y así cíclicamente.

Este procedimiento admite una implementación en tiempo casi real, pues existe un retardo constante equivalente al tiempo que dura la respuesta a impulso, más el tiempo de procesamiento del primer bloque. Si la respuesta tiene 128 muestras y la frecuencia de muestreo es de 44.1 kHz, el retardo mínimo es de unos 3 ms.

Cuando la longitud de la respuesta a impulso es pequeña, tanto la convolución lineal como la convolución por bloques (también denominada convolución rápida) son igualmente eficientes.

El método por bloques recién descrito es conocido en la literatura en idioma inglés como *overlap and add*. Una variante, *overlap and save*, superpone los bloques de entrada en lugar de los bloques de salida, descarta las primeras $n_h - 1$ muestras de cada bloque, y encadena los bloques resultantes sin necesidad de sumas. Si bien produce una mayor economía de procesamiento, tampoco funciona en tiempo real en el sentido estricto, dado que también genera un cierto retraso entre la entrada y la salida. Gardner (1995) propuso un algoritmo híbrido de convolución en tiempo real, basado en la combinación del método lineal y por bloques (*minimum delay block convolution*), en el que la convolución lineal produce resultados mientras se recolectan los datos para el procesamiento en bloques.

La síntesis binaural consiste, entonces, en la convolución de las formas de onda de un sonido a espacializar y de una respuesta a impulso registrada con una cabeza artificial dentro de una cámara anecoica. Si la respuesta a impulso fue tomada a 90° de desplazamiento horizontal – considerando 0° a la posición frontal- y con una elevación de 30° , el resultado de la convolución nos permite escuchar al sonido original como si proviniera de esa posición. Cambiando dinámicamente las respuestas a impulso podemos describir trayectorias en un espacio de tres dimensiones.

3.2 Implementación de un primer programa

Desarrollé un prototipo de localización espacial del sonido basado en síntesis binaural, en dos instancias. La primera fue programada en lenguaje “C”, para sistema operativo Windows de 32 bits, en la cual, el procesamiento de los archivos de audio se produce en tiempo diferido. Esta versión simula el comportamiento del sonido directo, sin considerar reflexiones, como si la audición se produjese al aire libre.

Para la segunda versión, creé un objeto para Max-MSP, programado en lenguaje “C”. Este objeto puede combinarse con otros, especialmente diseñados, para extender la base del modelo de acuerdo a las necesidades del usuario y disponibilidad de capacidad de proceso de la computadora utilizada.

3.2.0 Acerca de las HRTF utilizadas

Las respuestas a impulso usadas en ambos programas pertenecen al conjunto elaborado por Keith Martin y Bill Gardner, quienes produjeron una serie de mediciones sobre una cabeza artificial y las pusieron a disposición de la comunidad, en un proyecto del MIT Media Lab (Gardner y Martin, 1994).

Para las mediciones emplearon un maniquí KEMAR modelo DB-4004, con pabellones auriculares artificiales DB-061 para el lado izquierdo y DB-065 para el derecho, y micrófonos Etymotic ER 11. El sistema fue montado en una base móvil controlada por una computadora, en la cámara anecoica del MIT. Los impulsos fueron reproducidos a través de un parlante Realistic Optimus Pro 7 a una distancia de 1,4 metros. Se tomaron muestras desde 710 posiciones, con ángulos de elevación que van desde -40° hasta 90° y ángulos de azimut entre 0° y 360° . Como los pabellones artificiales utilizados eran diferentes, se tomaron mediciones sobre los 360° , obteniéndose así dos conjuntos completos de HRTFs simétricas.

La tabla siguiente muestra la cantidad de mediciones tomadas horizontalmente para cada ángulo de elevación, y su incremento, en grados.

Elevación °	Nº de mediciones	Incremento de azimut °
-40	56	6.43
-30	60	6.00
-20	72	5.00
-10	72	5.00
0	72	5.00
10	72	5.00
20	72	5.00
30	60	6.00
40	56	6.43
50	45	8.00
60	36	10.00
70	24	15.00
80	12	30.00
90	1	-

También se tomaron mediciones de la respuesta del parlante y de varios modelos de auriculares dispuestos sobre la cabeza artificial. La respuesta a impulso del parlante Optimus fue medida en la cámara anecoica usando un micrófono Neumann K Mi 8, a una distancia de 1,4 metros. Esta respuesta fue usada para crear un filtro inverso aplicable a las HRTFs.

Para las mediciones se usó una computadora Macintosh equipada con un conversor Audiomedia II. Las muestras fueron grabadas con una resolución de 16 bits y una frecuencia de muestreo de 44,1 kHz.

Los impulsos fueron generados con secuencias pseudo aleatorias de longitud máxima (ML), siguiendo la técnica de Rife y Vanderkoy (1989). Cada archivo tenía inicialmente 16383 muestras, y luego fue truncado a 512. Posteriormente, crearon un conjunto simétrico con 128 muestras por archivo, ecualizadas para compensar la respuesta no uniforme del parlante Optimus. Esta versión compacta fue pensada para aquellos interesados en la síntesis en 3D, y es la que utilizamos en este proyecto.

3.2.1 Preparación de las HRTF

Par el tratamiento de las HRTF escribí un programa en lenguaje “C”, cuyas funciones son las siguientes:

- abrir los 368 archivos estéreo que contienen las respuestas a impulso (HRIR) y ajustar su formato.
- agregar 128 ceros a continuación, preparándolos para la convolución rápida.
- realizar la FFT de cada respuesta a impulso
- ubicar los archivos FFT resultantes en orden sucesivo en un único archivo

3.2.2 Principios de funcionamiento del software

La figura 3.4 muestra el esquema básico de síntesis binaural aplicado en las versiones mencionadas. La señal de entrada ingresa por bloques de 128 muestras –la misma longitud que las respuestas a impulso- y duplica su extensión con ceros dispuestos a continuación de la última muestra leída. Esta función se traslada al dominio de la frecuencia por medio de la FFT.

Una unidad selecciona la HRTF a utilizar de acuerdo a la posición de la fuente virtual. Para ahorrar tiempo de procesamiento, las respuestas a impulso de la cabeza artificial (HRIR por

Head Related Impulse Response) fueron previamente procesadas, y almacenadas como números complejos (pares de valores, el primero real, el segundo imaginario) luego de aplicarles la FFT. Esta unidad, además, debe seleccionar las HRTF considerando que sólo cubren 180°. Esto significa que para una posición de -90°, hay que tomar las HRTF medidas a 90° y cruzar ambos canales (la HRTF izquierda se convierte en derecha y viceversa).

Si la posición de la fuente cambia, una unidad de interpolación crea una cantidad determinada de HRTF intermedias entre la anteriormente usada y la próxima. Se trata de una interpolación lineal a nivel espectral, que persigue la eliminación de los clicks audibles que normalmente se producen al cambiar de respuesta a impulso durante el proceso de convolución. La elección de la cantidad de términos depende en parte de las características de la fuente. Para una banda de ruido un valor bajo, o incluso nulo, puede ser suficiente. Para un sonido puro, en cambio, es necesario elevar el número de términos considerablemente para no generar discontinuidades severas en la forma de onda.

Sin embargo, es preciso considerar que mientras más alto es el número de términos, más baja es la precisión en la localización. La interpolación requiere un tiempo de realización, y si en medio de ese proceso la posición cambia nuevamente, el espectro calculado hasta ese momento se torna nuevo punto de partida hacia el nuevo espectro a considerar. Los cambios rápidos de posición, y una elevada cantidad de términos de la interpolación, pueden conducir a la percepción de una imagen confusa.

Posteriormente, se produce la multiplicación de las señales en el dominio de la frecuencia. Se trata de una multiplicación de los resultados de las FFT, expresados en forma de números complejos.

$$(a, b).(c, d) = (ac - bd, ad + bc) \quad (3.3)$$

Si denominamos a a la parte real del resultado, y b a la parte imaginaria, la amplitud y la fase de cada muestra del espectro está dada por:

$$A = \sqrt{a^2 + b^2} \quad (3.4)$$

$$\theta = \tan^{-1} \frac{b}{a} \quad (3.5)$$

Finalmente, el resultado de la multiplicación de la FFT de la señal de entrada por la HRTF correspondiente al oído izquierdo, y la HRTF correspondiente al oído derecho son transformadas al dominio del tiempo por la FFT inversa.

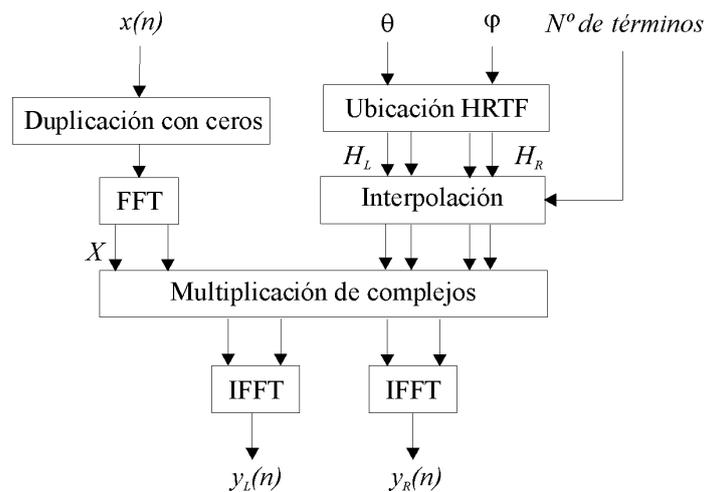


Figura 3.4 Esquema básico de funcionamiento del localizador por síntesis binaural.

El modelo, hasta aquí, sólo considera el posicionamiento de la fuente sobre una esfera imaginaria que rodea a la cabeza. Como se habrá observado, todavía nos falta implementar otras características antes mencionadas, como la distancia fuente sujeto, la acción filtrante del aire, el efecto del ambiente virtual, etc.

3.2.3 Versión para OS Windows

Una interfaz gráfica de operación simple (ver figura 3.5) permite especificar el ángulo de azimut, el ángulo de elevación y la distancia a partir de envolventes, en las cuales el eje del tiempo se adapta a la duración del archivo a procesar. La distancia es simulada a través de la caída en amplitud con la distancia, y la acción filtrante del aire en función de la frecuencia.

En el cuadro central se representa una proyección de la trayectoria en tres dimensiones, con la posibilidad de rotar el gráfico para una mejor apreciación. Por debajo, se encuentra el control para la definición de la cantidad de términos a considerar en la interpolación. Un valor igual a cero anula este procedimiento.

Desde los menús se accede a los archivos de entrada, se renombra el archivo de salida (que por defecto se almacena en un archivo temporal, de nombre *temp.wav*), y se guardan o recuperan archivos de trayectoria.

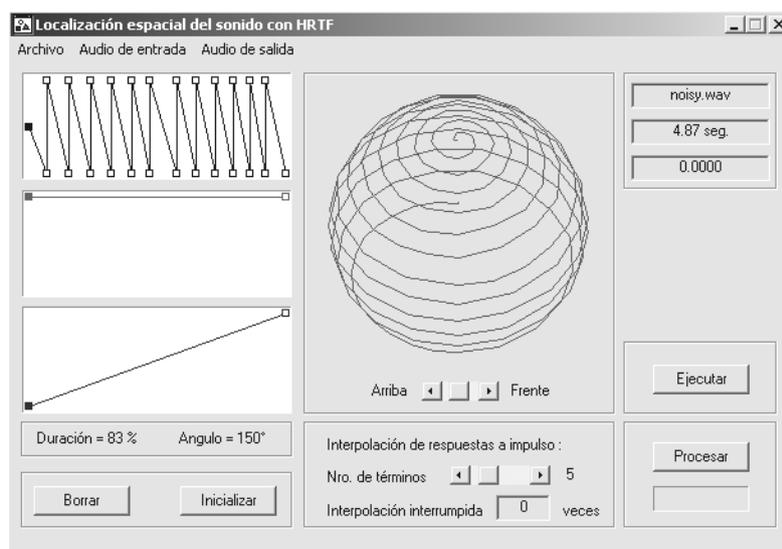


Figura 3.5 Software de localización espacial basado en síntesis binaural.

Versión para sistema operativo Windows de 32 bits.

3.3 Un sistema de localización espacial en tiempo real

3.3.0 Síntesis binaural con Max-MSP

Trasladé el código utilizado en la versión Windows a un objeto de Max-MSP, optimizando la unidad de búsqueda de las HRTF, y eliminando el control de la distancia fuente-sujeto con el propósito de independizar esta propiedad del objeto de síntesis binaural.

La figura 3.6 muestra el archivo de ayuda diseñado para este nuevo objeto, denominado *zhrtf~*. Los *inlets* admiten, de izquierda a derecha, la señal de entrada, el ángulo de azimut, el ángulo de elevación y la cantidad de términos de la interpolación entre HRTFs.

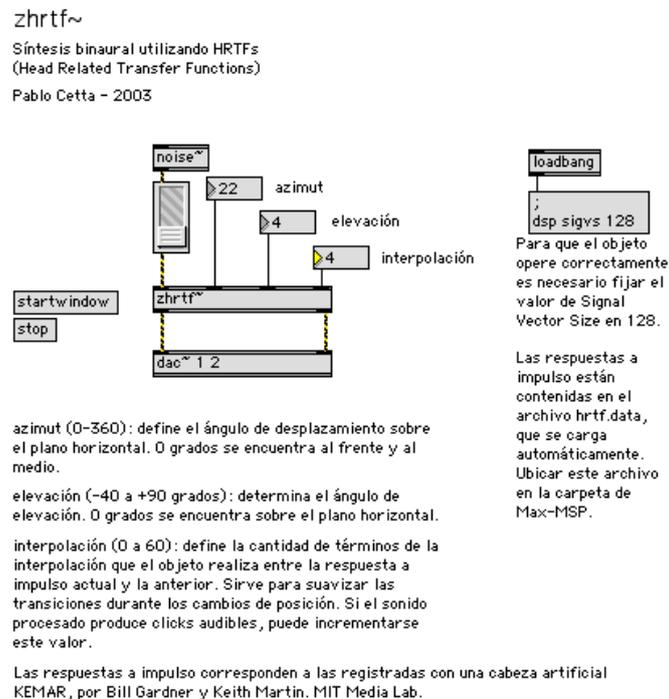


Figura 3.6 Archivo de ayuda del objeto de síntesis binaural, programado para el entorno Max-MSP.

3.3.1 Simulación de la absorción del aire

El objeto *airabsorb~* simula la absorción del aire en función de la distancia. La figura 3.7 muestra el contenido del archivo de ayuda de este objeto, que admite la señal a procesar, la distancia fuente-sujeto en metros, y un coeficiente subjetivo de absorción en sus entradas. En el ejemplo, la amplitud de la señal es escalada en función de la distancia, siguiendo la ley del cubo inverso.

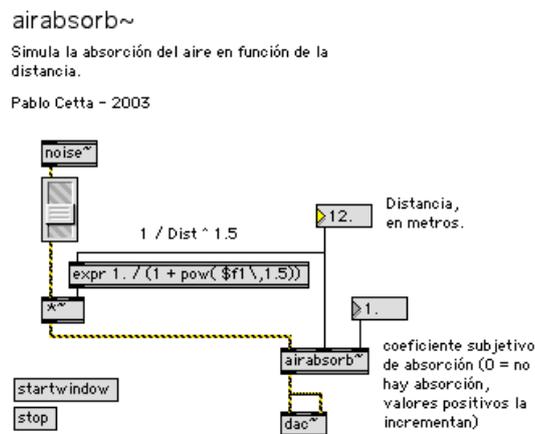


Figura 3.7 Objeto de simulación de la absorción del aire, programado para el entorno Max-MSP.

3.3.2 Simulación de las primeras reflexiones

El objeto *coord_phantom~* calcula la ubicación de las fuentes virtuales, responsables de la simulación de las seis primeras reflexiones en un recinto cerrado. Los *inlets* reciben las coordenadas cartesianas de la posición de la fuente en el espacio, y las dimensiones del ambiente, ancho, largo y alto, respectivamente. Las coordenadas de las fuentes fantasma (x , y , z) de las paredes derecha, izquierda, frente y atrás, el techo y el piso, son enviadas a los *outlets*, siguiendo ese orden, siempre de izquierda a derecha. En la figura 3.8 se puede apreciar el contenido del archivo de ayuda de este objeto.

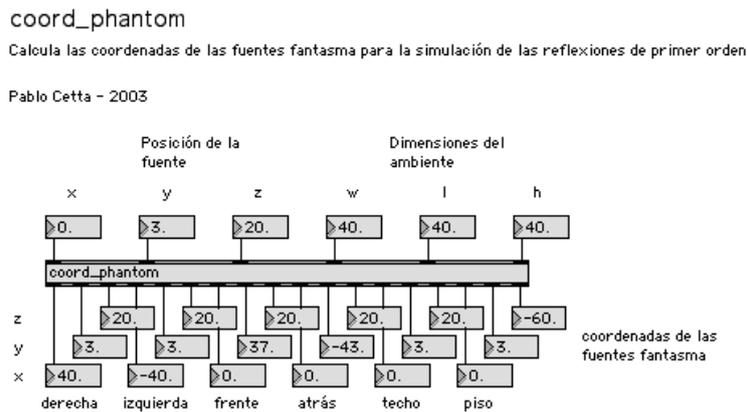


Figura 3.8 Pantalla de ayuda del objeto que calcula la posición de las fuentes fantasma, destinadas a la simulación de las primeras reflexiones.

El objeto *echo_del_gain~* (figura 3.9) recibe las coordenadas de una fuente fantasma y calcula la distancia entre esa posición y la ubicación del oyente. A partir del valor obtenido, aplica la caída en amplitud, el retardo, y el filtrado que simula la absorción del aire en relación a esa distancia, a una señal de audio que ingresa previamente a un objeto *tapin~*,

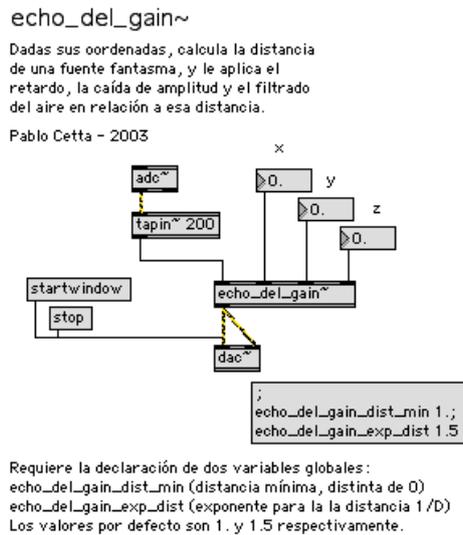


Figura 3.9 Pantalla de ayuda del objeto que calcula la distancia de una fuente fantasma, y aplica transformaciones a la señal de audio, en relación a esa distancia.

El objeto *dir_echoes~* recibe una señal de audio, la posición de la fuente que la produce y las dimensiones del ambiente virtual, y genera copias procesadas que simulan los seis primeros ecos. Junto con estas señales, reporta la ubicación de las fuentes fantasma que las generan, para indicar a las unidades de espacialización la dirección y distancia recorrida por cada reflexión. La figura 3.10 muestra la pantalla de ayuda de este objeto.

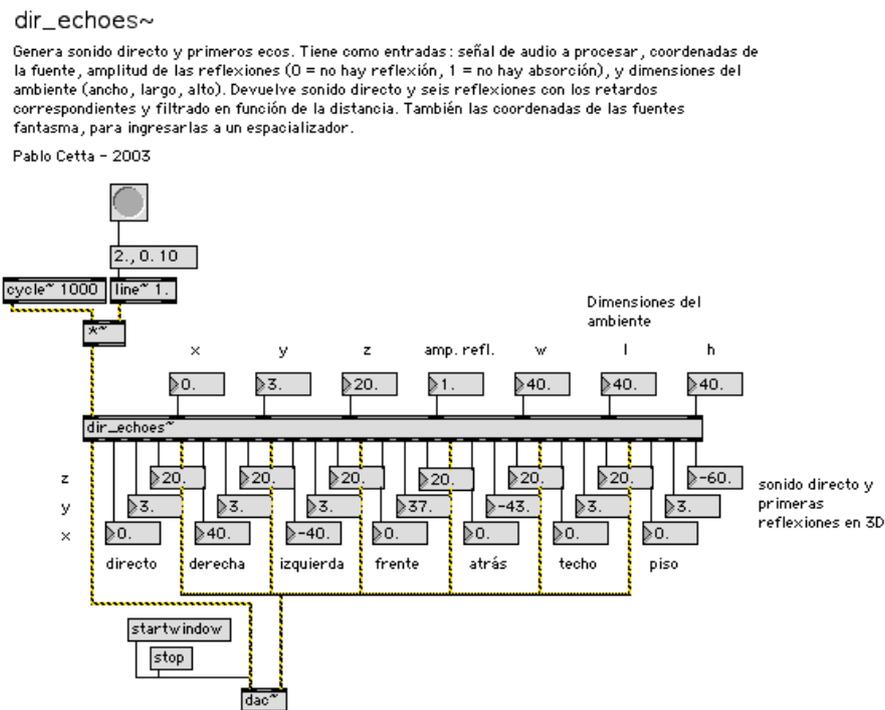


Figura 3.10 Pantalla de ayuda del objeto que procesa el sonido directo y las primeras reflexiones, dadas la posición de la fuente y las dimensiones del ambiente.

La inclusión de un modelo acústico para la simulación de los primeros ecos y la reverberación contribuye a lograr un efecto de mayor realismo (Begault, 1991, o Hartmann y Wittemberg, 1996). Esta idea encuentra su justificación, en principio, si consideramos que la audición de fuentes reales en ambientes anecoicos resulta antinatural. La discriminación de la posición de la fuente puede verse fortalecida, particularmente en aquellos casos donde el sonido posee un ataque lento o cuando el ataque es enmascarado. No obstante, la sensación de realismo no

siempre está emparentada con una mayor precisión en la discriminación, debido a los efectos de localización aditiva y precedencia.

3.3.3 Descripción de la posición de la fuente

Considerando las aplicaciones musicales que veremos más adelante, resulta necesario disponer de un sistema de ubicación de las fuentes virtuales de posibilidades múltiples. En este sistema, la posición de una fuente se define a través de dos vectores. El primero tiene su origen en la posición del oyente, y puede trasladarse y rotar. El segundo tiene el origen en el extremo del primer vector, y sus movimientos de traslación y rotación son siempre referidos a ese origen relativo (figura 3.11).

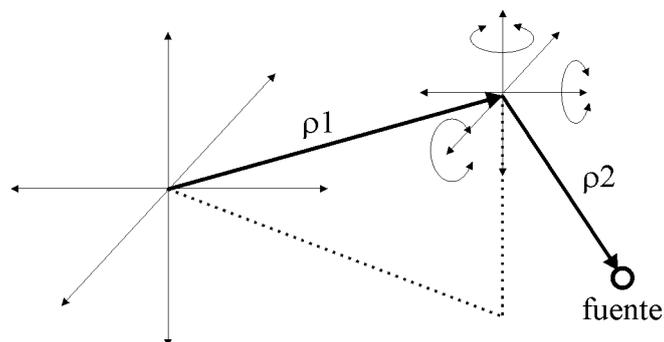


Figura 3.11 Posicionamiento de la fuente a través de dos vectores solidarios.

Para precisar la posición de un punto P en el espacio partimos de un sistema de coordenadas cartesianas ortogonales, cuyo origen se encuentra en la intersección de los ejes, e identificamos la posición a través de sus componentes, $P = (x, y, z)$.

También podemos definir la posición del punto en el espacio a través de coordenadas esféricas, donde ρ es la distancia entre el punto P y el origen O , θ es el ángulo de

lateralización, y φ el ángulo de elevación. Las coordenadas cartesianas del punto, definido por sus coordenadas esféricas son:

$$\begin{aligned}x &= \rho \cos \varphi \sin \theta \\y &= \rho \cos \varphi \cos \theta \\z &= \rho \sin \varphi\end{aligned}\tag{3.6}$$

Si hacemos variar θ entre 0° y 360° , y a φ entre -90° y $+90^\circ$, el vector \vec{OP} define todos los puntos del espacio que se encuentran sobre una esfera de centro O , y radio ρ .

Resulta a veces conveniente expresar algunas trayectorias en coordenadas cartesianas, y otras en coordenadas a esféricas. Por tal motivo, programé dos objetos en “C” para estos fines. Las pantallas de ayuda de Max-MSP se muestran en la figura 3.12.

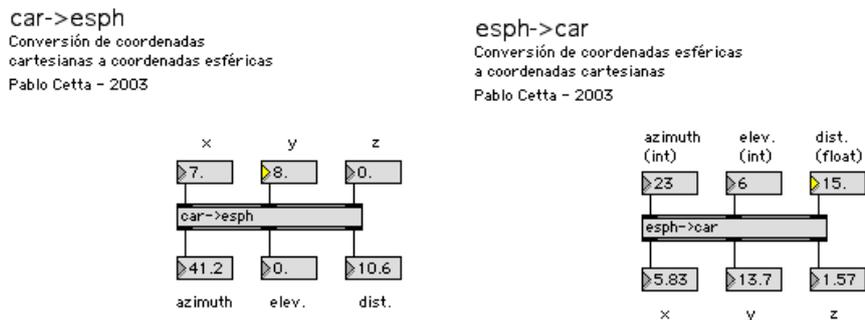


Figura 3.12 Objetos que convierten coordenadas cartesianas a esféricas, y viceversa.

3.3.4 Aplicación de transformaciones geométricas

Rotación alrededor de los ejes

La rotación alrededor del eje Y sólo provoca la modificación en las componentes x y z del punto P . Las tres ecuaciones siguientes nos permiten conocer las coordenadas de un punto después de haberlo rotado alrededor del eje Y , un ángulo ω_y .

$$\begin{aligned}
 x' &= x \cos \omega_y - z \sin \omega_y \\
 y' &= y \\
 z' &= z \cos \omega_y + x \sin \omega_y
 \end{aligned}
 \tag{3.7}$$

Este mismo sistema podemos escribirlo como un sistema de ecuaciones lineales y encolumnarlo, sumando en cada ecuación la variable ausente multiplicada por cero.

$$\begin{aligned}
 x' &= x \cos \omega_y + y \cdot 0 - z \sin \omega_y \\
 y' &= x \cdot 0 + y + z \cdot 0 \\
 z' &= x \sin \omega_y + y \cdot 0 + z \cos \omega_y
 \end{aligned}
 \tag{3.8}$$

La rotación sobre el eje X está dada por:

$$\begin{aligned}
 x' &= x + 0 y + 0 z \\
 y' &= 0 x + y \cos \omega_x + z \sin \omega_x \\
 z' &= 0 x - y \sin \omega_x + z \cos \omega_x
 \end{aligned}
 \tag{3.9}$$

Y la rotación alrededor del eje Z por:

$$\begin{aligned}
 x' &= x \cos \omega_z + y \sin \omega_z + 0 z \\
 y' &= -x \sin \omega_z + y \cos \omega_z + 0 z \\
 z' &= 0 x + 0 y + z
 \end{aligned}
 \tag{3.10}$$

Traslación

El sistema de ecuaciones que nos da las coordenadas de un punto después de aplicarle una traslación es:

$$\begin{aligned}
 x' &= x + \Delta x \\
 y' &= y + \Delta y \\
 z' &= z + \Delta z
 \end{aligned}
 \tag{3.11}$$

si lo escribimos de modo que aparezcan todas las variables en cada ecuación tendremos:

$$\begin{aligned}x' &= x + 0 y + 0 z + \Delta x \\y' &= 0 x + y + 0 z + \Delta y \\z' &= 0 x + 0 y + z + \Delta z\end{aligned}\tag{3.12}$$

A diferencia de los sistemas que indican la posición del punto después de rotarlo, éste presenta en cada ecuación un sumando que no está afectado por ninguna de las variables. Frente a una necesidad que se nos presentará más adelante, veremos otra forma de escribir este sistema, a través de un cambio de coordenadas.

Utilización de coordenadas homogéneas

Las coordenadas homogéneas permiten referir un punto conocido en un sistema de referencia de n dimensiones a otro de $n + 1$ dimensiones.

Si conocemos $P = (x, y, z)$ podemos expresarlo en coordenadas homogéneas como $P' = (x', y', z', M)$ siendo $M \neq 0$, $x' = Mx$, $y' = My$, $z' = Mz$, entonces $P' = (Mx, My, Mz, M)$. Conocido P' en coordenadas homogéneas podremos reescribirlo en coordenadas cartesianas donde $x = x'/M$, $y = y'/M$, $z = z'/M$. Entonces para cada punto $P = (x, y, z)$ existen infinitas coordenadas homogéneas, una por cada $M \neq 0$.

Nos interesa el caso particular en el que $M = 1$, ya que esta elección no provoca alteraciones entre las coordenadas conocidas del punto en el sistema homogéneo con respecto al sistema no homogéneo.

Como no es posible graficar esta situación para un punto en 3D, dado que en coordenadas homogéneas se expresaría en cuatro dimensiones, explicaremos este caso particular del cambio a coordenadas homogéneas partiendo de un punto referido a un sistema bidimensional.

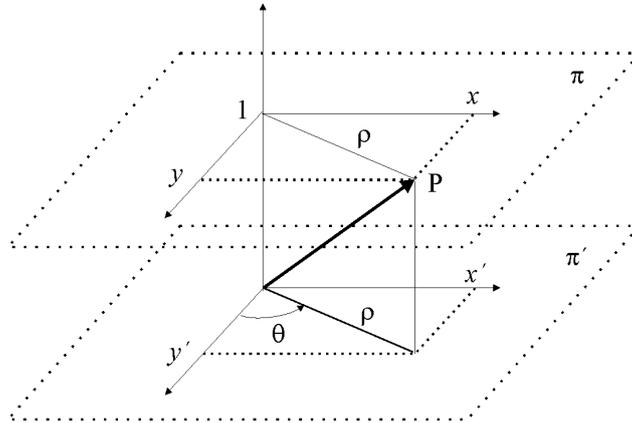


Figura 3.13 Coordenadas homogéneas en 2D, para $M = 1$.

Supongamos un plano π , sobre el que ubicamos un punto P . Sobre el plano podemos dibujar un sistema de referencia, con ejes Y y X . Las coordenadas del punto $P = (x, y)$ pueden obtenerse a partir de la proyección del radio vector que lo posiciona, sobre los ejes de coordenadas:

$$\begin{aligned} x &= \rho \sin \theta \\ y &= \rho \cos \theta \end{aligned} \quad (3.13)$$

Imaginemos este plano π paralelo a otro plano, π' , que se encuentra a una distancia igual a una unidad de él. El punto P contenido en el plano π , puede estar ahora referido a un sistema de tres coordenadas, $x, y, 1$. Entonces, $P' = (x', y', 1)$, cuyas componentes son:

$$\begin{aligned} x' &= \rho \sin \theta \\ y' &= \rho \cos \theta \\ 1 &= 1 \end{aligned} \quad (3.14)$$

Las coordenadas x' e y' no se alteraron con respecto al sistema de referencia original, es decir, $x' = x$, $y' = y$. La tercera ecuación $1 = 1$ se denomina trivial, y no influye en la determinación de las coordenadas del punto respecto del sistema original, ni tampoco en sus movimientos.

Retornemos a nuestra situación original, frente a la necesidad de expresar la traslación del punto P a través de un sistema de ecuaciones lineales, vemos que los incrementos que determinan la traslación no aparecen multiplicados por ninguna variable. Pero si a P lo expresamos en coordenadas homogéneas, con $M = 1$, resultará $P = (x, y, z, 1)$ con:

$$\begin{aligned}x &= \rho \cos \varphi \sin \theta \\y &= \rho \cos \varphi \cos \theta \\z &= \rho \sin \varphi \\1 &= 1\end{aligned}\tag{3.15}$$

y el sistema que indica la traslación de P podría entonces escribirse:

$$\begin{aligned}x' &= x + 0 y + 0 z + 1 \Delta x \\y' &= 0 x + y + 0 z + 1 \Delta y \\z' &= 0 x + 0 y + z + 1 \Delta z \\1 &= 0 x + 0 y + 0 z + 1\end{aligned}\tag{3.16}$$

Hemos utilizado a 1 como cuarta variable, que multiplica a los incrementos. Haciendo extensivo el principio a la rotación sobre el eje Y , por ejemplo, tenemos:

$$\begin{aligned}x' &= \cos \omega_y x + 0 y - \sin \omega_y z + 0 1 \\y' &= 0 x + y + 0 z + 0 1 \\z' &= \sin \omega_y x + 0 y + \cos \omega_y z + 0 1 \\1 &= 0 x + 0 y + 0 z + 1 1\end{aligned}\tag{3.17}$$

Podemos, entonces, expresar cada uno de los movimientos a través de sistemas lineales de cuatro ecuaciones y cuatro variables. En consecuencia, la rotación de un punto alrededor del eje Y puede expresarse como un producto de matrices. La primera de ellas se forma encolumnando los coeficientes de cada variable, y la segunda, encolumnando las variables mismas:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \omega_y & 0 & -\sin \omega_y & 0 \\ 0 & 1 & 0 & 0 \\ \sin \omega_y & 0 & \cos \omega_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.18)$$

La matriz de los coeficientes se denomina *matriz de rotación* alrededor del eje Y (R_y). Podemos llamar X' a la matriz de las incógnitas y X a la matriz de las variables, por lo cual, la rotación sobre Y se define por:

$$X' = R_y X \quad (3.19)$$

Del mismo modo podemos definir matrices de transformación para la rotación alrededor de los ejes Z y X , y para la traslación:

$$R_z = \begin{bmatrix} \cos \omega_z & \sin \omega_z & 0 & 0 \\ -\sin \omega_z & \cos \omega_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \omega_x & \sin \omega_x & 0 \\ 0 & -\sin \omega_x & \cos \omega_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Si deseamos aplicar una traslación, luego una rotación alrededor del eje Y , a continuación una rotación alrededor del eje Z , y finalmente una rotación alrededor del eje X , son necesarias transformaciones sucesivas:

$$X' = \{ [R_x (R_z R_y)] T \} X \quad (3.20)$$

Pero debe tenerse en cuenta que no es lo mismo trasladar un punto y luego rotarlo, que el camino inverso, dado que el producto de matrices para cada transformación no es el mismo.

Es usual en la práctica transponer las matrices de las incógnitas (X') y de las variables (X), convirtiendo las matrices columna en matrices fila. La razón de esta transposición se justifica en el modo de representación de las matrices en los lenguajes de programación.

La transpuesta de un producto de matrices es igual al producto de las matrices transpuestas en orden inverso, por ejemplo, si $A = B C$ entonces $A^T = (BC)^T = C^T B^T$. La ecuación 3.20, por lo tanto, puede escribirse como:

$$(X')^T = X^T \{ T^T [Ry^T (Rz^T Rx^T)] \} \quad (3.21)$$

Dada esta propiedad del álgebra de matrices, y que en programación se presentan las matrices directamente transpuestas, será usual leer: $X' = X \{ T [Ry (Rz Rx)] \}$.

3.3.5 Desarrollo de una interfaz gráfica

Programé el sistema de la figura 3.11 como un objeto de Max-MSP, escrito en lenguaje “C”, recurriendo al álgebra matricial para calcular la posición de la fuente en coordenadas cartesianas, de acuerdo a las características de los dos vectores. El extremo del primer vector se ubica a través de las coordenadas rectangulares que ingresan en los tres primeros *inlets*, aunque es posible definir su posición mediante coordenadas esféricas y convertirlas con el objeto antes mencionado. Esto puede ser útil cuando el vector debe rotar, por ejemplo. Los tres *inlets* siguientes admiten ángulos de rotación entre 0 y 360°, que son aplicados a los ejes del sistema relativo de coordenadas, origen del segundo vector. Los tres últimos *inlets* definen la posición del extremo del segundo vector en forma cartesiana. La figura 3.14 muestra la pantalla de ayuda de este objeto de control, denominado *coord*.

coord

Calcula las coordenadas de un punto (fuente sonora) cuya posición está determinada por dos vectores. Uno rota sobre sus ejes (rojo), mientras que el otro aplica una traslación (azul). La fuente se encuentra en el extremo del vector rojo.

Pablo Cetta - 2003

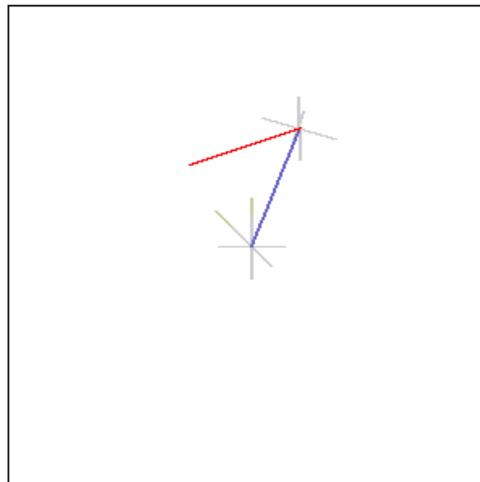
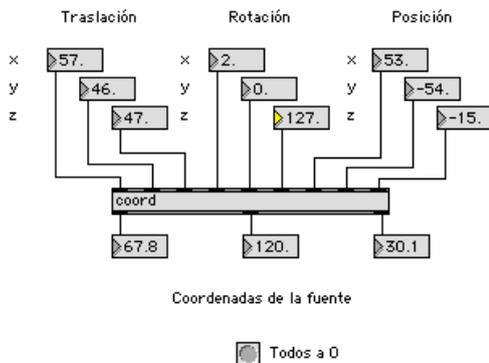


Figura 3.14 Objeto de control *coord*, que calcula la posición final de la fuente al dar datos sobre los vectores que definen su posición.

El gráfico de la derecha, en la figura anterior, es creado con otro objeto que muestra el movimiento de los vectores y de la fuente en tiempo real. Este objeto se denomina *tr_graph*, y su archivo de ayuda se muestra en la figura 3.15.

tr_graph

Grafica las coordenadas de un punto (fuente sonora) cuya posición está determinada por dos vectores. Uno rota sobre sus ejes (rojo), mientras que el otro aplica una traslación (azul). La fuente se encuentra en el extremo del vector rojo.

Pablo Cetta - 2003

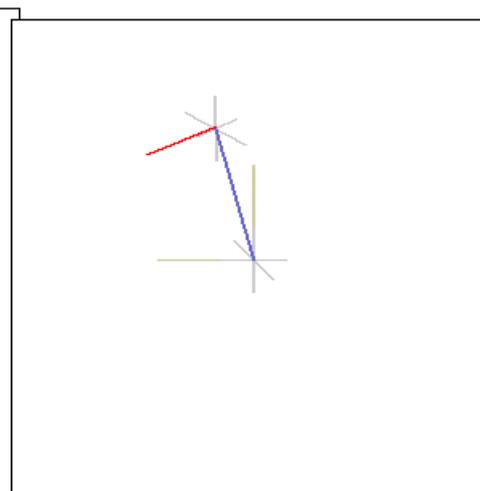
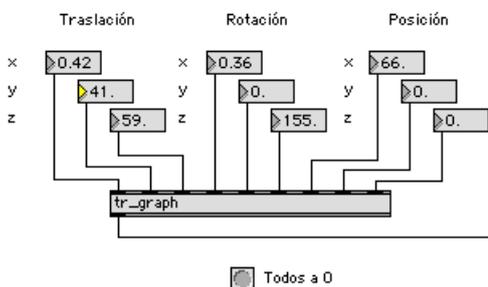


Figura 3.15 Pantalla de ayuda del objeto de visualización de los movimientos de la fuente *tr_graph*.

Los objetos mencionados en esta sección se han integrado en una interfaz gráfica, que sirve de herramienta para el trazado de trayectoria o posiciones fijas de las fuentes virtuales. Los parámetros pueden variar dinámicamente por medio de envolventes, que expresan posiciones tanto en coordenadas cartesianas como esféricas. La figura 3.16 muestra la pantalla principal de este subprograma.

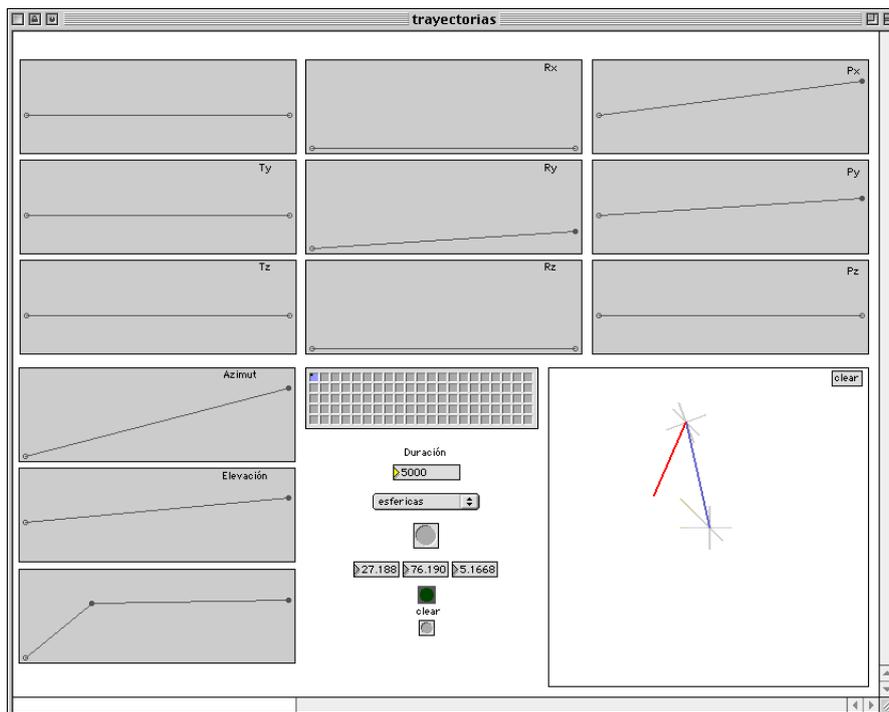


Figura 3.16 Pantalla principal de la interfaz gráfica para el trazado de trayectorias.

Por último, diseñé un objeto capaz de tomar muestras de una trayectoria, y descomponerlas sobre los tres ejes en gráficos independientes. No sólo permite almacenar trayectorias y modificarlas a voluntad por medios gráficos, sino que los resultados almacenados pueden ser aplicados como control de otros parámetros sonoros (figura 3.17).

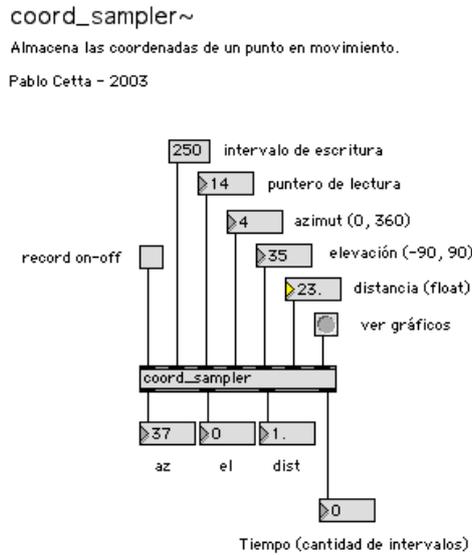


Figura 3.17 Objeto destinado al muestreo y almacenamiento de trayectorias espaciales.

3.3.6 Reproducción transaural

Para la reproducción de señales grabadas binauralmente, o tratadas con síntesis binaural, suelen emplearse auriculares, que aseguran que la señal destinada a cada uno de los oídos ingresa directamente al conducto auditivo, lugar en el que fueron ubicados los micrófonos al grabar con la cabeza artificial. La reproducción con parlantes es, en cambio, contaminada por las reflexiones naturales del ambiente en el que se reproduce la grabación, por la doble acción del oído externo (registrada al grabar y puesta nuevamente en acción al escuchar), por las reflexiones en torno a la cabeza y el torso, y fundamentalmente, por el cruzamiento de información que se produce entre los dos oídos. La señal destinada al oído izquierdo parte del parlante izquierdo e ingresa a ese oído, pero también al derecho, y viceversa. Esto trae como consecuencia una deformación significativa de la imagen espacial, debida a la percepción confusa de las diferencias interaurales. Este efecto de cruzamiento –denominado en inglés *crosstalk*, por analogía con ciertos fenómenos de la telefonía– puede ser disminuido significativamente empleando filtros. Este problema no es exclusivo de las técnicas binaurales, sino que afecta a todos los sistemas basados en la reproducción con parlantes.

Básicamente, se busca que cada parlante emita la señal destinada al oído correspondiente, pero sumada a otra que anule el efecto no deseado respecto al oído contrario. A la reproducción binaural con parlantes se la denomina reproducción *transaural*.

La teoría de la cancelación del *crosstalk* fue tratada en principio por Bauert (1961), y luego por Schroeder y Atal (1963). Posteriormente, Cooper y Bauck (1989) reelaboraron la teoría anterior en base a un método simplificado destinado a la cancelación del *crosstalk*, y a la generación de parlantes virtuales en posiciones arbitrarias.

El proceso de eliminación del *crosstalk* ideado por Schroeder puede representarse a través de la siguiente matriz:

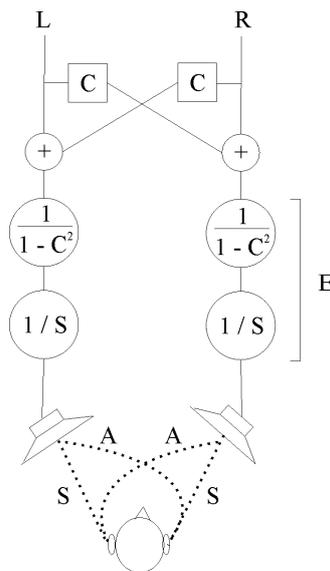


Figura 3.18 Matriz para eliminación del *crosstalk* propuesta por Schroeder (1963).

$C(z)$ y $E(z)$ son filtros de cancelación y ecualización, respectivamente.

Conocemos de antemano las respuestas a impulso $s(n)$ y $a(n)$, que pueden obtenerse grabando con una cabeza artificial un impulso que parte del mismo ángulo en el que se ubican los parlantes (30° , por ejemplo). $s(n)$ corresponde al oído más cercano y $a(n)$ al más alejado.

Vamos a determinar la función de transferencia de los filtros C y E en términos de $S(z)$ y $A(z)$, las transformadas z de los impulsos. A efectos de simplificar la notación vamos a denominar a estas funciones sólo por la letra que las identifica (C, E, S, A)

De la observación de la matriz de la figura 3.18 deducimos lo siguiente:

$$\begin{aligned} R' &= R [E S + C E A] + L [E A + C E S] \\ R' &= [R [S + C A] + L [A + C S]] E \end{aligned} \quad (3.23)$$

Para que la eliminación del *crosstalk* se haga efectiva, es necesario que R' dependa sólo de R (señal destinada al oído derecho) y no de L (señal destinada al oído izquierdo). Por lo tanto, para que esto suceda, es preciso que lo que multiplica a L sea igual a cero:

$$A + C S = 0$$

Por lo tanto,

$$C = -\frac{A}{S} \quad (3.24)$$

Por otra parte, se requiere que R' sea igual a R , para que no exista ninguna transformación de la señal luego de haberse eliminado el *crosstalk*. Por esto, y por el hecho de haber ubicado los parlantes en frente del oyente en lugar de ubicarlos contra sus oídos, es necesario emplear el filtro de ecualización E .

Para que R' sea igual a R ,

$$[S + C A] E = 1$$

Reemplazando C ,

$$\left[S + \left(-\frac{A}{S} \right) A \right] E = 1$$

$$E = \frac{1}{S + \left(-\frac{A}{S}\right)A}$$

Multiplicando ambos términos por S nos queda:

$$E = \frac{S}{S^2 - A^2} = \frac{S}{S^2 \left(1 - \frac{A^2}{S^2}\right)}$$

$$E = \frac{1}{S} \frac{1}{(1 - C^2)} = E_1 \cdot E_2 \quad (3.25)$$

Obtuvimos dos filtros de ecualización $E_1 = 1/S$, destinado a corregir la deformación introducida por la ubicación de los parlantes respecto al oyente y $E_2 = 1/(1 - C^2)$ que corrige la deformación generada a partir de la eliminación del *crosstalk*.

Cooper y Bauck dispusieron el modelo de Schroeder en una matriz de filtros equivalente (figura 3.19 a), en la cual

$$A' = -\frac{A}{S^2 - A^2} \quad \text{y} \quad S' = \frac{S}{S^2 - A^2} \quad (3.26)$$

y la confrontaron con otra versión optimizada, obteniendo los mismos resultados. La nueva matriz (denominada *shuffler*, figura 3.19 b) se reduce a la implementación de dos filtros definidos por

$$N' = \frac{1}{S - A} \quad \text{y} \quad P' = \frac{1}{S + A} \quad (3.27)$$

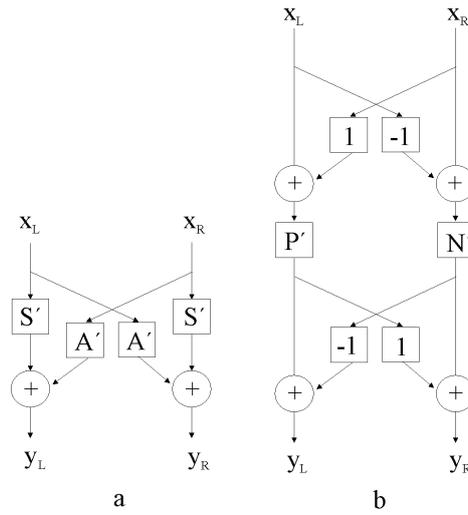


Figura 3.19 Matriz *lattice* (a) y matriz *shuffler* (b) comparadas.

La figura 3.20 muestra una implementación de la matriz *shuffler* para reproducción transaural, que supone una disposición simétrica de los parlantes

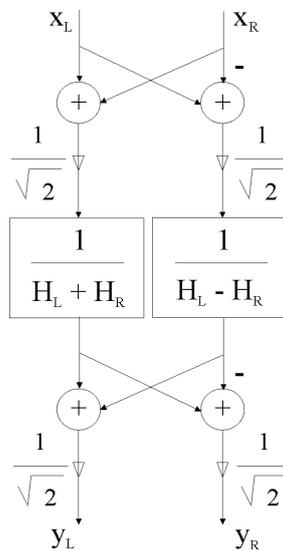


Figura 3.20 Implementación de matriz *shuffler* para reproducción transaural.

La figura 3.21 muestra el objeto sobre el cual implementé el módulo de eliminación del *crosstalk* y ecualización de las señales. Las señales binaurales ingresan por los dos primeros *inlets*, mientras que los dos restantes reciben el ángulo de azimut y elevación de uno de los parlantes (valor positivo entre 0° y 180°). La posición del otro parlante se obtiene de la posición del primero, descartando que se trata de una disposición simétrica respecto al oyente.

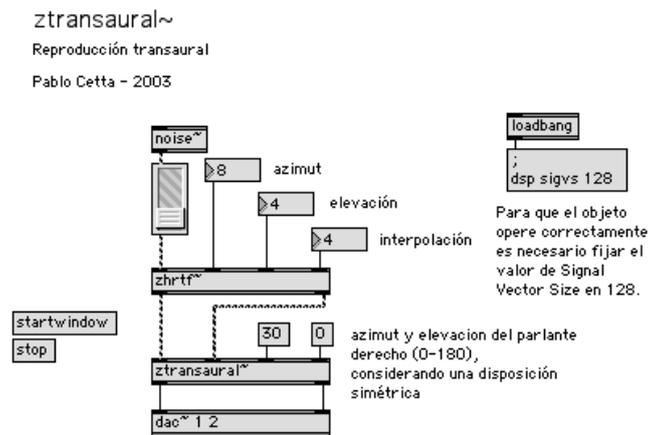


Figura 3.21 Pantalla de ayuda del objeto *ztransaural*, para reproducción transaural.

En la reproducción transaural, las reflexiones naturales que se producen en el espacio de audición resultan críticas. Kendall y Martens (1984) observaron la formación de imágenes espaciales en un tipo de ambiente controlado, al cual Davis y Davis (1980) denominaron LEDE (*live end, dead end*). Este ambiente es utilizado en las salas de monitoreo de algunos estudios de grabación, y consta de superficies absorbentes al frente y reverberantes en la parte posterior. La deformación de la imagen de acuerdo a la posición del oyente puede apreciarse en la figura 3.22.

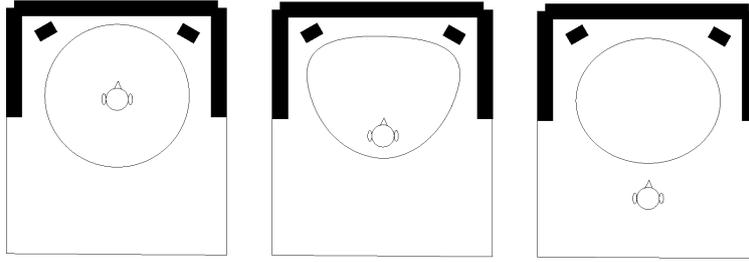


Figura 3.22 Deformación de la imagen transaural en un ambiente LEDE, según la posición del oyente (Kendall y Martens, 1984).

Para preservar la imagen espacial resulta necesario un control del ambiente, que impida que las reflexiones naturales arriben antes de los 20 ms de producido el sonido directo. En ambientes no controlados, resulta prácticamente imposible producir imágenes detrás de la cabeza con los parlantes ubicados al frente. En estos casos, la cuadrafonía convencional parece más adecuada.

3.3.7 Extensión a la cuadrafonía

Con el propósito de eliminar las ambigüedades frente-atrás que se producen en la detección de la posición de una fuente aparente, propias de la técnica binaural, realicé una extensión del sistema transaural al ámbito de la cuadrafonía. El principio se basa en la realización de un *crossfade* entre las señales destinadas a los parlantes delanteros y los traseros, en función del ángulo de lateralización de la fuente. Un objeto, denominado *zpan_trans~*, recibe las señales binaurales, y el ángulo de azimut que define la posición de la fuente, y las distribuye a dos unidades de reproducción transaural, de forma similar a lo que ocurre en la simulación por *intensity panning*. En los objetos *ztransaural~* se define el ángulo de azimut, y el ángulo de elevación, de los parlantes anterior y posterior derecho. Esta configuración se muestra en la figura 3.23.

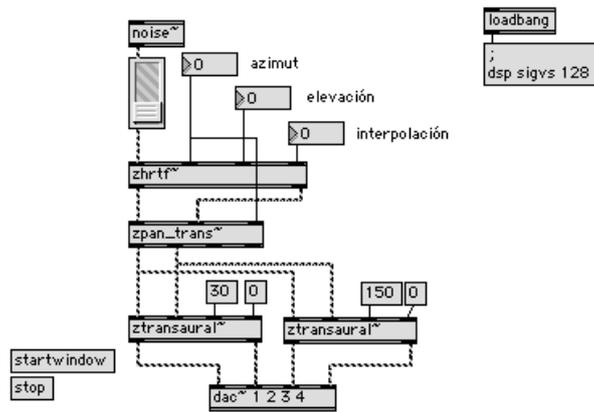


Figura 3.23 Extensión de la reproducción transaural a la cuadrafonía, a través del objeto *zpan_trans*.

A través de un mecanismo similar, es posible extender este principio a la reproducción octofónica, siempre considerando que la demanda de capacidad de procesamiento disminuye las posibilidades del entorno compositivo.

3.3.8 Configuraciones posibles

Los objetos creados admiten cierta variedad en cuanto a posibilidades de interconexión. En un mismo entorno es posible utilizar más de una técnica de espacialización del sonido, e incluso compartir objetos destinados a la simulación de la absorción del aire, del efecto Doppler, o de las primeras reflexiones y la reverberación. La técnica doble transaural cuadrafónica, por ejemplo, aplicada a un plano sonoro principal, puede coexistir con Ambisonics decodificado para cuatro parlantes, aplicado a planos secundarios dispuestos horizontalmente. Si bien el segundo sistema ofrece posibilidades de simulación sólo en dos dimensiones, presenta un consumo más bajo de los recursos de la computadora.

3.4 Integración de la música al espacio virtual

Los eventos sonoros, al igual que las estructuras musicales, se desenvuelven en un espacio propio e interno, delimitado por los parámetros que los caracterizan. Algunas propiedades de ese espacio son observables a través de los movimientos de un intérprete o un director de orquesta, que transforma y exterioriza ciertos recorridos inscriptos en la partitura, en gestos visibles, o lo que es lo mismo, en movimientos plenos de significación.

La localización espacial del sonido aplicada a la composición, siguiendo esta idea, es concebida como una exteriorización del espacio interno de la música, y a la vez, como una interiorización del espacio externo. Significa que las trayectorias espaciales se desenvuelven en relación a los recorridos interiores del sonido, y que los parámetros del sonido o la música son alterados por la trayectoria de las fuentes virtuales.

Consideramos, entonces, dos aspectos:

- a) El sonido genera la trayectoria.
- b) La trayectoria transforma al sonido.

El primer caso involucra la detección de la magnitud de algunos parámetros elegidos, como la altura, la dinámica, el grado de tonicidad o la duración de los sonidos instrumentales que ingresan a una computadora. Estos parámetros se agrupan por ternas, y a cada uno de ellos se le asigna una componente del espacio. Si los parámetros varían, cambian en consecuencia las coordenadas de la posición de la fuente. Las acciones musicales que realizan los instrumentistas provienen de la partitura, donde se contempla la modificación gradual de estos parámetros, no sólo en función del discurso a generar sino también de los recorridos espaciales.

La detección de los comportamientos sonoros provenientes de los instrumentos se realiza por medio de objetos especialmente creados, como el que se representa en la figura 3.24, destinado a registrar los cambios de altura del sonido, cuando la amplitud supera un cierto umbral preestablecido.

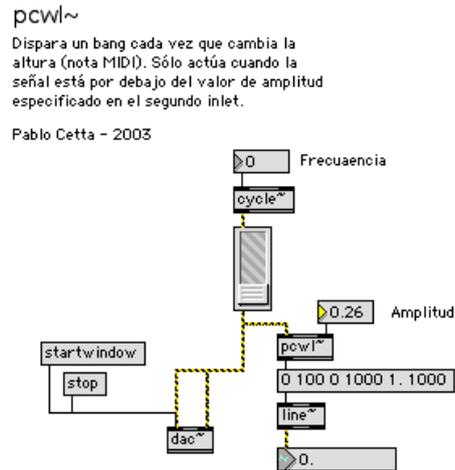


Figura 3.24 Objeto que detecta cambios de altura, y los reporta, sólo si ocurren por encima de un cierto umbral de amplitud.

En el segundo caso, la descomposición de las trayectorias sobre sus ejes, producen valores que alteran la magnitud de los parámetros de procesamiento de las señales provenientes de los instrumentos.

3.4.0 Exteriorización e interiorización del espacio

Los procesos comparativos definen distancias en el espacio interno de la música, que se relacionan con posiciones fijas del espacio virtual. Los procesos modulatorios, en cambio, ocurren en el tiempo y definen una trayectoria específica en ese espacio interno, que se traduce en un movimiento de las fuentes sonoras.

La aplicación más inmediata de la exteriorización del espacio interno es la registración espacial de la altura. Cualquier disposición puede ser considerada, pero vamos a mencionar sólo una, donde a cada grado del total cromático corresponde un ángulo de azimut, y a cada octava un ángulo de elevación. Esta disposición coincide con cierta forma de representar la altura, a través de una espiral ascendente, en la cual todos los grados de igual nombre coinciden verticalmente.

Si disponemos los eventos sonoros secuencialmente, el sentido de giro de la fuente se relaciona con la direccionalidad melódica, y la velocidad depende de la interválica presente. Si los intervalos son cada vez más amplios, la velocidad aumenta en consecuencia. Frente a la superposición de estratos melódicos podemos hablar, entonces, de un contrapunto extendido que involucra a las trayectorias.

Desde el punto de vista de la implementación, debemos considerar que el recorrido de la fuente debe presentar curvas suavizadas en el tiempo. Esto lo logramos a través de rampas de cierta duración entre valores sucesivos de altura, que amortiguan las desviaciones de los giros melódicos. También es preciso considerar el nivel de tonicidad del sonido, a fin de decidir la posición de la fuente en caso que no exista altura definida.

Con el propósito de evitar que el oyente “aprenda” con excesiva rapidez las relaciones entre altura y espacio virtual, realizamos transformaciones de la posición de la espiral, tales como desplazamientos o rotaciones, como parte de las modificaciones ambientales.

Otro aspecto a considerar es el relativo a las distancias. La simulación de la distancia debe depender en primer grado de la intensidad de los eventos sonoros. Mientras menor es la intensidad, más lejos se encuentra la fuente del eje de la espiral. Para incrementar el realismo de la situación, implementamos un efecto filtrante, asociado a la amplitud de la señal. De este modo, las dinámicas tenues se asocian a espectros pobres y las fuertes a espectros más

brillantes. La figura siguiente muestra un esquema de la unidad donde se implementó este procesador.

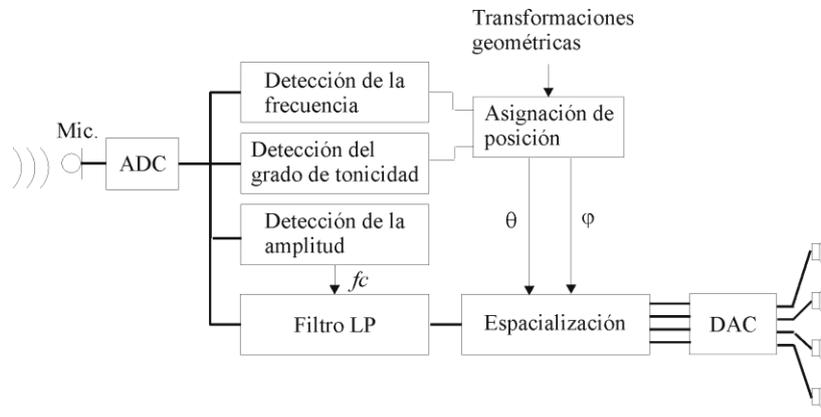


Figura 3.25 Esquema de implementación de módulo de espacialización en función de la altura, la dinámica y el grado de tonicidad de los eventos sonoros.

La figura 3.26 muestra un ejemplo donde el registro, el ámbito melódico y la dinámica de una sucesión de sonidos, conforman las componentes de una trayectoria espacial, que actúa como correlato del movimiento interno del gesto musical (a). La misma trayectoria sirve para transformar los parámetros de un proceso de filtrado aplicado a otro objeto sonoro (b). Los parámetros de la sucesión melódica, traducidos al ámbito de la música electrónica, se convierten ahora en frecuencia de corte, ancho de banda y ganancia del filtro.

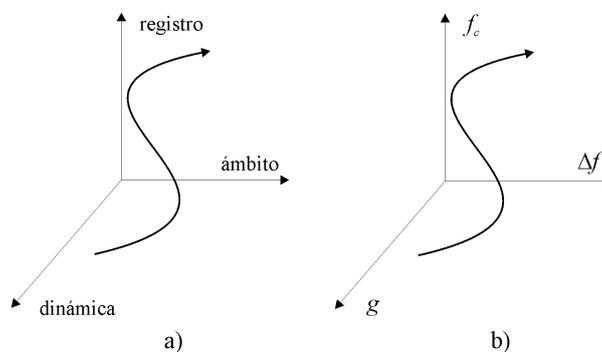


Figura 3.26 Relación entre los parámetros de un gesto musical y su desplazamiento en el espacio (a) y utilización de la misma trayectoria en el procesamiento de un objeto sonoro con un filtro (b).

Extendiendo este principio a otros parámetros, podemos crear múltiples trayectorias para una misma fuente –tantas como ternas de variables consideremos-, o bien, aplicar las coordenadas definidas por la trayectoria de una fuente al control de la magnitud de los parámetros de otro sonido o fragmento de música. De este modo, la evolución dinámica de un evento sonoro, su tonicidad, el nivel de consonancia de las redes interválicas de un sector musical, su densidad cronométrica o polifónica, o la relación de periodicidad o aperiodicidad de una sucesión de eventos, se constituyen como variables del espacio, o parámetros transformados por las trayectorias en él descriptas.

También es posible intercambiar los parámetros asignados a cada eje de coordenadas del espacio virtual a fin de rotarlo, o modificar el factor de escala, a la manera de un *zoom* espacial. Las coordenadas obtenidas pueden sufrir transformaciones geométricas -generadas por el objeto *coord* antes descrito-, y las mutaciones del sonido o las secuencias de eventos pueden coincidir en tiempo con los movimientos de la fuente, o bien ser anticipadas, o retardadas a la manera de un eco. La magnitud de los parámetros, traducidas tanto a coordenadas cartesianas como esféricas, duplica las trayectorias posibles, sin que por esto se altere la percepción de estas relaciones. Por último, cualquier parámetro del sonido o la música puede afectar no sólo a las trayectorias, sino también a las dimensiones o características acústicas del espacio virtual.

3.4.1 Tratamiento espacial de la altura y el timbre

La organización de conjuntos de grados cromáticos (PCS por *pitch class sets*), como método de control de la altura, resulta eficaz. He dedicado cierto tiempo al estudio de sus propiedades, y al desarrollo de programas de composición asistida que utilicé luego en la composición de algunas obras (Cetta, 2003b).

Del análisis de las características de los conjuntos, surgen clasificaciones orientadas a la generación de similitudes, oposiciones o transformaciones entre campos de altura. En el caso de los conjuntos formados por cuatro sonidos, por ejemplo, una clasificación posible surge en relación al grado de asociación respecto al sistema tonal -acordes de séptima, tríadas con notas agregadas, acorde bimodal, conjuntos cuyos subconjuntos son acordes por terceras, acordes por cuartas, fragmentos de escalas reconocibles- o disociación respecto a este sistema, ordenados por sus características propias.

Las relaciones entre los campos de altura se regulan en función al nivel de consonancia de cada conjunto. La determinación del grado de consonancia puede establecerse por varios métodos, pero básicamente tiene en cuenta la cantidad y calidad de las clases interválicas presentes en cada conjunto, y la disposición vertical u horizontal de sus sonidos en el registro. En el ordenamiento se comparan todos los conjuntos, pero también se establecen subcategorías de acuerdo a ausencia de uno o más intervalos en particular, cantidad de clases interválicas presentes, cantidad de transposiciones o inversiones no redundantes, etc.

Cuando varios instrumentos monódicos ejecutan simultáneamente las notas de un conjunto, cada sonido se diferencia por su timbre, por el modo de ejecución empleado y por la ubicación de cada instrumentista en el escenario. Cualquiera de estas características puede ser transformada, aplicando el procesamiento adecuado. Pero es posible, además, simular la producción de varias alturas organizadas con un único instrumento monódico, transportando aquella que ejecuta a otras distancias, y aplicando a cada nota un tratamiento tímbrico diferenciado y una ubicación particular en el espacio.

Para la implementación de este proceso, programé un objeto de control que genera los índices de transposición a utilizar en la transposición (ver figura 3.27). Estos índices son calculados para conjuntos de grados cromáticos de tres y cuatro elementos, cuya denominación se especifica en uno de sus *inlets*. Si elegimos el conjunto 4-15, por ejemplo, e informamos al

programa que la nota ejecutada corresponde a la posición del bajo, al tocar un *do*, es posible obtener *do#*, *mi* y *fa#*, que completan el conjunto antes mencionado. El objeto calcula una transposición posible y decide si lo invierte o no, por lo cual, en cada interpretación de la misma nota escuchamos un acorde con la misma estructura interválica (4-15, en el ejemplo) pero transpuesta, invertida, o permutada de formas distintas.

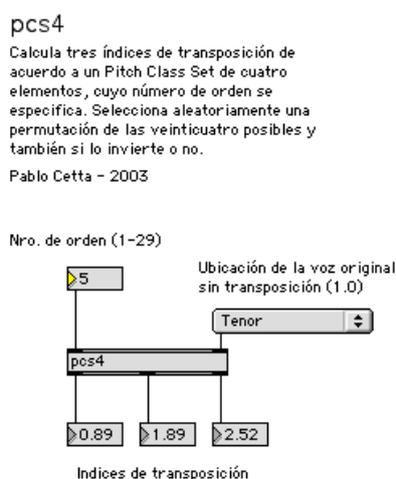


Figura 3.27 Objeto que calcula índices de transposición de acuerdo a un PCS especificado.

Las transposiciones del sonido de un mismo instrumento presentan un nivel de coherencia mucho mayor que en el caso de varios instrumentos ejecutando las notas de un acorde. Obviamente la relación de las componentes espectrales es la misma para todas, pero desplazada en frecuencia. Esta característica genera una integración de los sonidos, que guía a la percepción de un único espectro, por lo cual, este tratamiento de la altura produce resultados más bien tímbricos. El efecto se potencia en la combinación de estos acordes-espectros con el tratamiento interválico de las fundamentales.

Con el propósito de generar acciones musicales variadas, recurrimos a un objeto que detecta si la nota supera en duración un cierto umbral. Sólo en ese caso la operación de armonización es llevada a cabo. Esto permite la ejecución de giros rápidos –adornos, por ejemplo- sin

armonización y detenciones, donde la armonía aparece a través de un *crescendo* programado. La distribución espacial de estas acciones complementa el efecto buscado.

Por otra parte, mediante el procesamiento adecuado, el sonido de un instrumento puede convertirse en el de otro, y la modulación entre ambos ser acompañada por una trayectoria, que los vincula espacialmente.

3.4.2 Tratamiento espacial de las relaciones temporales

Es posible comparar el tiempo subjetivo de cada fragmento musical, o sus modulaciones, con las dimensiones del espacio virtual. Desde un punto de vista perceptual, al aumentar el espacio, la sensación de transcurso del tiempo parece disminuir. Esta idea, considerada a través del tema que nos ocupa, puede justificarse a través de diversas causas, una es la acción de la reverberación, que disminuye la resolución temporal. Otra surge en relación a los primeros ecos; las diferencias temporales entre los tiempos de arribo de las primeras reflexiones (ver figura 3.28) aumentan a medida que el espacio crece.

Dejemos de lado, por el momento, la disminución en la intensidad del sonido debida a la distancia, y la absorción asociada a la reflexión, aprovechando la virtualidad de nuestras aplicaciones del espacio. Si las dimensiones del ambiente virtual crecen considerablemente a medida que transcurre el tiempo, llega un momento en que las primeras reflexiones son percibidas con un grado de separación tal, que se pueden expresar en términos aplicables al ritmo.

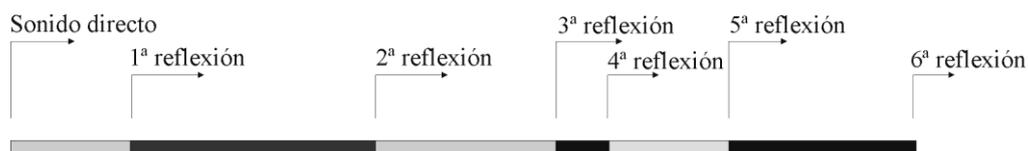


Figura 3.28 Relaciones entre los tiempos de llegada de las primeras reflexiones, medidos con segmentos.

Si modificamos la dirección de la fuente, o la distancia, percibimos un cambio en la estructura temporal generada por las reflexiones. La figura 3.29 muestra cómo se transforma esa estructura, cuando el ángulo de azimut varía entre 0° y 45° . El gráfico surge del cálculo del tiempo de arribo de las seis primeras reflexiones, para una elevación de la fuente equivalente a 5° , y una distancia igual a 0.483 veces el lado del recinto, formado por un cubo.

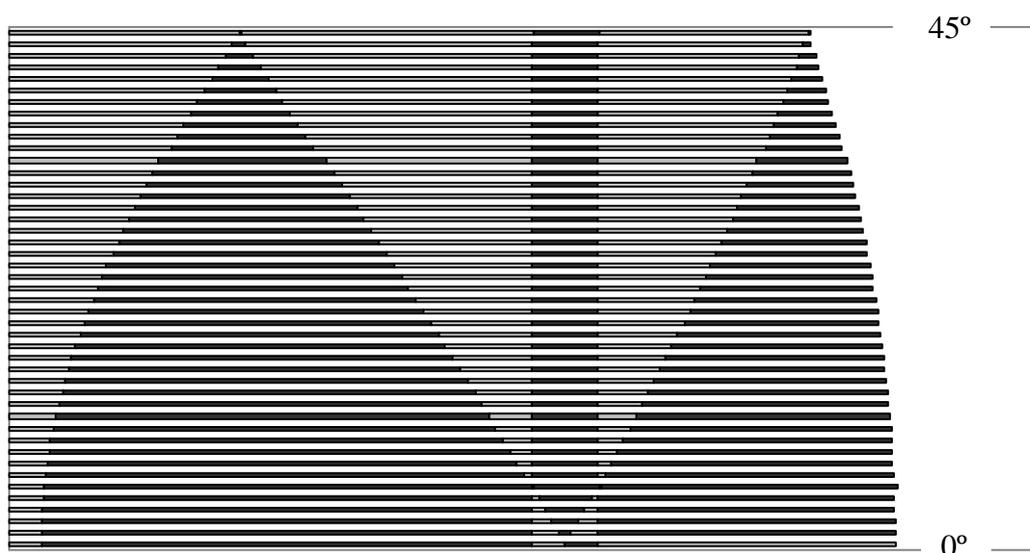


Figura 3.29 Relaciones entre los tiempos de llegada de las primeras reflexiones, variando el ángulo de azimut.

En este sentido, las dimensiones del ambiente virtual funcionan como un indicador de *tempo* de las reflexiones. La figura siguiente es una traslación del gráfico anterior a notación musical, en el cual la duración más pequeña se expresa como apoyatura de la semicorchea. Una traslación similar es posible, usando corcheas de quintillo o seisillo, como figuras de menor valor.

Figure 3.30 displays a sequence of 40 measures of musical notation, numbered 0 through 39. The notation is presented in a vertical column. Each measure contains a series of notes and rests, often grouped by slurs or beams. The notation includes various rhythmic values such as quarter notes, eighth notes, and sixteenth notes, along with rests and dynamic markings. The sequence shows a progression of rhythmic patterns, with some measures featuring complex groupings and others being simpler. The notation is written in a standard musical staff format, with a treble clef and a key signature of one flat (B-flat).

Figura 3.30 Ídem figura 3.29, pero representado en notación musical.

Se aprecian claramente las características modulatorias que surgen de la disposición sucesiva de estas células. Los ritmos se generan del ataque de los instrumentos, pero además, surgen de

las acentuaciones, del ordenamiento perceptual de sucesiones de grados cromáticos, de la incorporación de adornos o articulaciones diferenciadas o de la modificación del timbre en cada evento.

Los ritmos son generados con un programa (ver figura 3.31), que utiliza los mismos objetos usados en el modelo de localización espacial del sonido propuesto, y que fueron destinados a la simulación de las primeras reflexiones. Este software toma las trayectorias diseñadas en la interfaz gráfica antes vista, y calcula los ritmos derivados de los primeros ecos. Los resultados pueden aplicarse no solo a la producción de esquemas rítmicos a ejecutar por los instrumentos, sino también al control temporal en tiempo real de otras unidades de procesamiento del sonido.

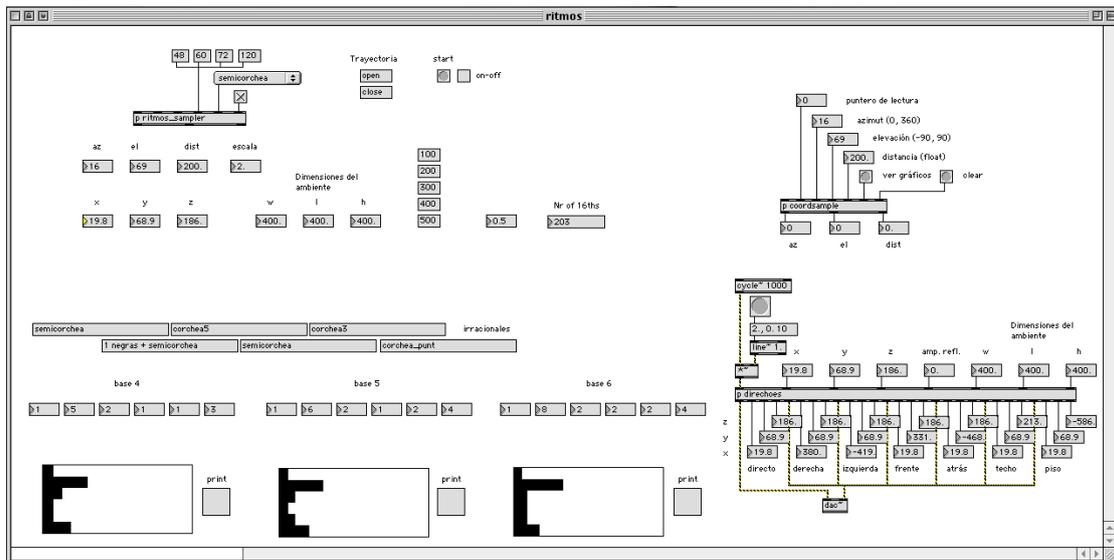


Figura 3.31 Software que genera ritmos a partir de trayectorias.

Hemos visto una vinculación perceptible de la relación espacio-tiempo. Sólo resta resaltar sus coincidencias, o bien ponerlas en conflicto, mediante el control de las dimensiones espaciales

y las densidades temporales. La compresión y expansión del tiempo y el espacio, siempre en términos musicales, se combinan del siguiente modo:

TIEMPO	ESPACIO
Compresión	Compresión
Compresión	Expansión
Expansión	Compresión
Expansión	Expansión

La expansión del espacio, y los ritmos asociados a las primeras reflexiones, se vincula de algún modo con la dilatación temporal propia del espectralismo. En esta música, la dilatación del tiempo da lugar a la resíntesis instrumental de las componentes sinusoidales de un sonido analizado. De no aplicarse tal aumento a la grilla temporal, las complejas variaciones espectrales de las componentes que ocurren durante el ataque del sonido, no podrían ser imitadas con instrumentos mecánicos convencionales. El procesamiento electroacústico permite, en nuestro caso, un recorrido gradual que parte de las reflexiones ambientales de un sonido instrumental y conduce a la generación de ritmos de ataque de varios instrumentos distribuidos en el espacio.

3.4.3 Tratamiento espacial de la textura

Los sonidos impulsivos, según vimos antes, resultan más fáciles de localizar que los sonidos continuantes. Las acciones musicales formadas por sucesiones de sonidos breves, por analogía, brindan mayores posibilidades espaciales que otras construidas sobre sonidos largos. La figura siguiente, presentada a modo de ejemplo, reúne algunos comportamientos típicos.



Figura 3.32 Sucesiones de sonidos breves, agrupadas de acuerdo a cantidad de alturas ejecutadas.

Un modo artificial de crear sucesiones de sonidos impulsivos, resulta de la fragmentación aleatoria de sonidos continuantes. El objeto que se muestra en la figura 3.33 hace uso de este principio, con el propósito de atomizar un evento sonoro y distribuirlo en el espacio, en un área de dimensiones variables, definible por el usuario.

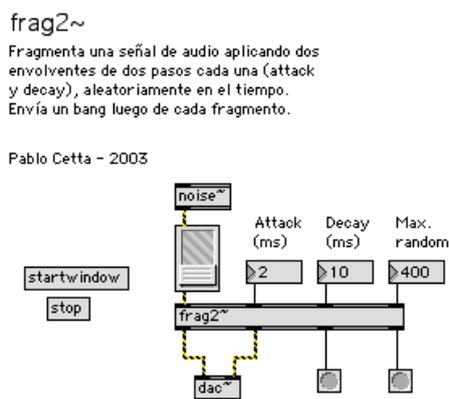


Figura 3.33 Objeto que fragmenta un evento continuante.

Para la distribución de los fragmentos en el espacio se recurre a otro objeto, que se muestra en la figura 3.34. El mismo genera coordenadas de forma aleatoria, comprendidas en una zona especificada, que puede modificarse en el tiempo, incluso con algún parámetro sonoro.

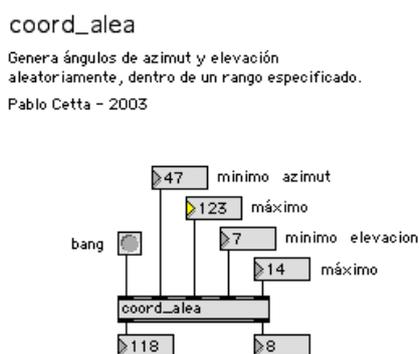


Figura 3.34 Objeto que devuelve coordenadas aleatorias dentro de un rango especificado.

Las herramientas de asistencia en la composición se presentan bajo la forma de módulos, que reciben las señales a procesar y los parámetros del procesamiento de forma remota (a través de variables). La figura 3.35, a modo de ejemplo, muestra el módulo que corresponde al

programa de fragmentación (a) y el generador de conjuntos de grados cromáticos por transposición antes mencionado (b).

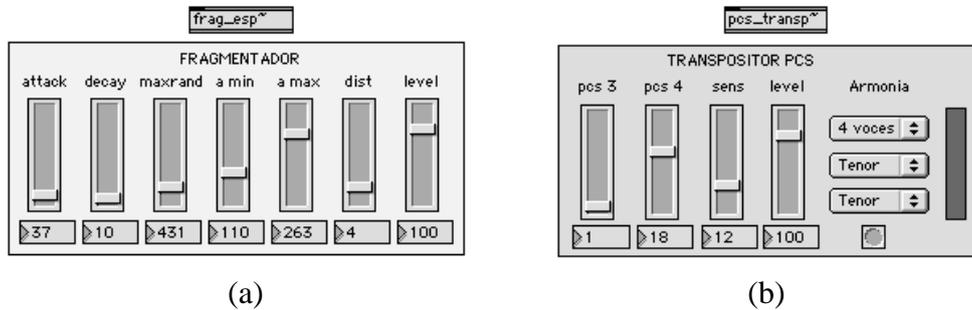


Figura 3.35 Módulos de fragmentación (a) y transposición según PCS (b).

Otra forma de tratamiento espacial de la textura consiste en la generación de volúmenes, delimitados por fuentes vinculadas, que reproducen eventos provenientes de instrumentos musicales, y sonidos sintetizados.

Las herramientas destinadas al control preciso de las coordenadas que definen la posición de las fuentes virtuales, vistas con anterioridad, fueron diseñadas con el propósito de facilitar la generación de movimientos complejos de varias fuentes que se desplazan en conjunto.

En la figura 3.36 observamos distintas formas geométricas, definidas por la ubicación de las fuentes aparentes, que rotan sobre su centro, o bien se trasladan en el espacio. Los números representan la cantidad de emisores involucrados en cada caso.

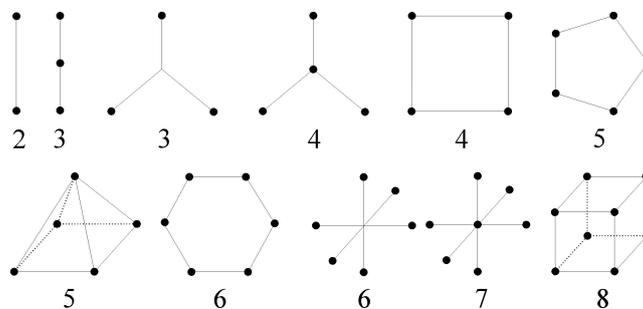


Figura 3.36 Ubicación de fuentes sonoras en complejos geométricos.

La textura global de estas formas en rotación se presenta como una sucesión de texturas sonoras o musicales individuales, emanadas por cada elemento del conjunto. El movimiento pone de relieve las dimensiones y la forma de estos complejos sonoros, compuestos principalmente de acciones iterativas, cuyos elementos se dispersan velozmente, y construyen luego otros volúmenes.

Podemos imaginar que los complejos permanecen fijos, mientras el oyente los rodea, como si se tratara de esculturas sonoras. El desplazamiento virtual del oyente alrededor de los objetos puede simularse mediante rotaciones y traslaciones del ambiente en su totalidad, manteniendo fijas las posiciones relativas de cada conjunto dentro del recinto.

Esto conduce a pensar un tipo de música, en la cual los recorridos en el espacio virtual son definidos por el mismo oyente, quien se detiene y examina cada objeto sonoro, rodeándolo y apreciando su textura según su voluntad. Algunos complejos permanecen fijos, y otros se desplazan sin intervención del que escucha, en un intrincado laberinto que invita a encontrar una salida.

En la figura 3.37 se reproduce la pantalla principal del programa diseñado para la obra *Interiores*. En la parte superior izquierda vemos los controles de entrada y salida de señales, y a continuación, el módulo de reverberación. En las filas siguientes, unidades de modulación en anillo, banco de filtros *comb*, el control de la simulación de las primeras reflexiones, fragmentadores, transpositores, generadores de volúmenes espaciales y la unidad de registración espacial de la altura. La matriz de la derecha y los controles inferiores permiten la interconexión de los módulos.

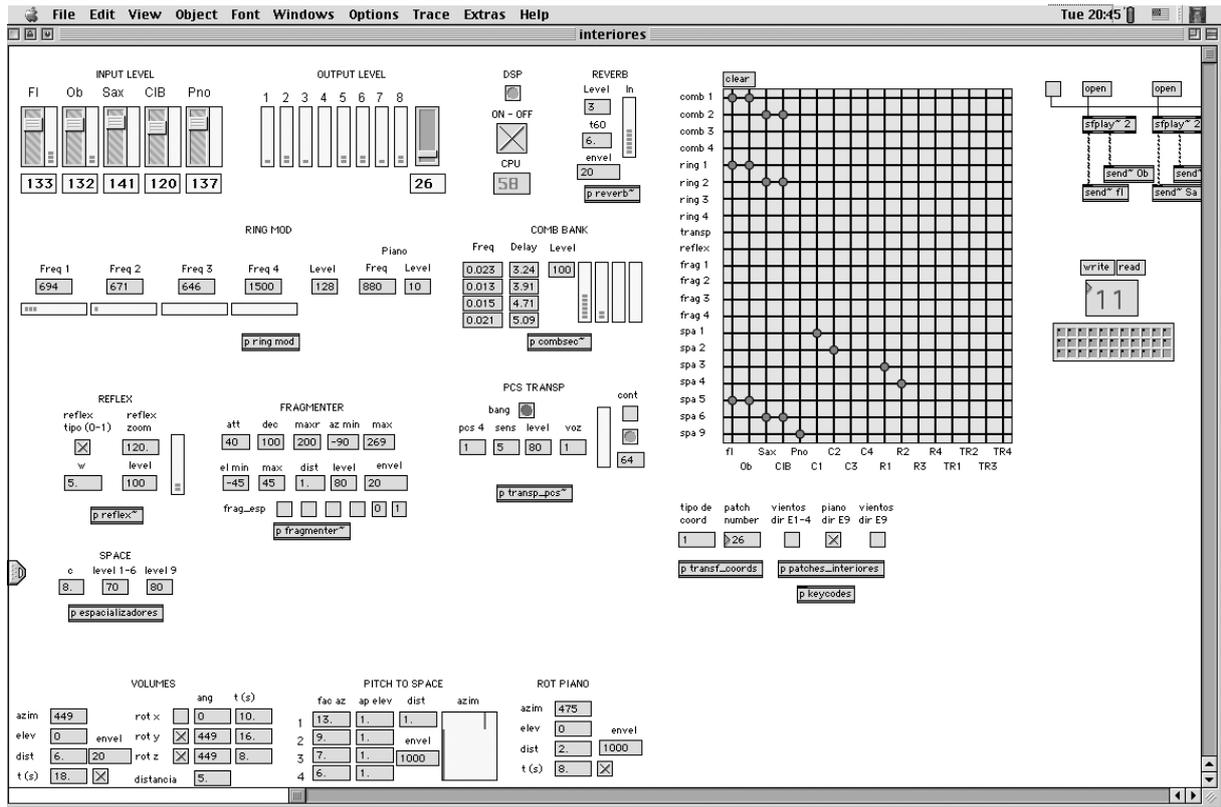


Figura 3.37 Pantalla principal del programa de procesamiento de *Interiores*.

A fin de facilitar la programación de las conexiones, se realizaron configuraciones preestablecidas que determinan el tipo de procesamiento de los instrumentos de viento y el piano. Los números en las tablas siguientes muestran los valores que el programa admite. A modo de ejemplo, si deseamos que los vientos sean afectados por transposición y filtros *comb*, y enviados a los fragmentadores, debemos cargar en el programa el número 107.

VIENTOS	Espacializadores	Fragmentadores	Simul. reflexiones.
Directo	1	101	201
Comb	2	102	202
Ring	3	103	203
Transp.	4	104	204
Ring – Comb	5	105	205
Comb – Ring	6	106	206
Transp. – Comb	7	107	207
Transp. – Ring	8	108	208
Transp. – Ring – Comb	9	109	209

PIANO	Espacializadores	Fragmentadores	Simul. reflexiones.
Directo	51	151	251
Comb	52	152	252
Ring	53	153	253
Transp.	54	154	254
Ring – Comb	55	155	255
Comb – Ring	56	156	256
Transp. – Comb	57	157	257
Transp. – Ring	58	158	258
Transp. – Ring – Comb	59	159	259

Para cada configuración debe indicarse, además, el tipo de distribución espacial a utilizar (ver tabla siguiente). El programa consta de siete espacializadores, uno de ellos dedicado especialmente al piano, y el resto en relación al tipo de procesamiento requerido.

Valor	Tipo de espacialización
1	Generación de volúmenes espaciales (una fuente en cada vértice de una figura geométrica, que puede rotar y desplazarse)
2	Registración espacial de la altura (cada “nota” de los vientos se ubica en una posición fija)
3	Generación de primeras reflexiones (o ritmos, al crecer exageradamente las dimensiones del ambiente virtual)
4	Distribución controlada de gránulos de fragmentación

A lo largo de la obra se emplean 31 *patches* automatizados, ordenados secuencialmente. Se avanza en la secuencia por medio del teclado de la computadora. Los cambios de *patch* se indican en la parte inferior de cada sistema de la partitura, con números enmarcados con rombos.

3.5 Descripción general de la obra

Introducción y Sector A. Comienzo con ruido eólico producido por el soplo en los instrumentos de viento, que luego es gradualmente coloreado por medio del banco de filtros *comb*. Para la espacialización se utiliza el generador de volúmenes espaciales a fin de crear un efecto envolvente. Se trata de una modulación en grado de tonicidad, comportamiento que se traslada a la parte de piano, pero con modulación en términos de consonancia-disonancia (conjuntos de grados cromáticos 4-1, 4-4, 4-12, 4-13, 4-16, 4-18 y 4-27, manteniendo el *do* como sonido común en el bajo). Hacia el final se aplica la fragmentación de los acordes tenidos.

Sector B. De gran libertad rítmica desde la escritura, ordenada por las acciones del director. Las articulaciones de los acordes, enmarcadas por el piano, son cada vez más complejas y variadas. Aumenta gradualmente la rugosidad de los acordes tenidos, partiendo de sonidos lisos hasta llegar a los trinos. La estructura de alturas es también modulante. El tipo de distribución espacial de los eventos se realiza por registración espacial de la altura, y luego por simulación de los primeros ecos, en un ambiente virtual que crece en dimensiones.

Sector C. Sirve de nexa a lo que sigue. Los trinos se desplazan por el espacio simulado con trayectorias amplias y a gran velocidad.

Sector D. Generación de volúmenes espaciales. Los vientos, con y sin procesamiento, se ubican en los vértices de una forma geométrica que rota sobre sus ejes, y a la vez se traslada. A fin de generar un tipo de textura localizable en términos de espacialización, los instrumentos producen giros de cuatro sonidos, repetidos a gran velocidad (por momentos también se aplica fragmentación). Las relaciones interválicas de los giros repetidos modulan, partiendo del conjunto de grados cromáticos 4-1, hasta llegar por orden sucesivo al 4-22. La rapidez de cada grupo se indica de acuerdo a tres niveles: $v-1$, $v0$ y $v1$. El piano, por otra

parte, se desenvuelve en un plano independiente, pero al mismo tiempo articula los ataques de los vientos. La escritura de esta sección manifiesta cierta libertad desde el punto de vista temporal. La unión efectiva de las partes depende en gran medida de la interpretación del piano y del director.

Sector E. Se divide en dos partes. La primera – sonidos largos- da lugar a la fragmentación como proceso. La segunda explora la percepción de las primeras reflexiones.

Sectores F, H, J y L. Solos a cargo de los vientos. Las relaciones interválicas de la melodía se trasladan a la armonización de las notas largas, realizadas con el transpositor por conjuntos de grados cromáticos, a cuatro voces. También se aplican otros procesos (*ring mod.*, filtros, fragmentación) y el tipo de espacialización es sutil y diferenciada para cada voz de los acordes generados.

Sectores G, I y K. Interludios instrumentales que unen los solos. Cada uno con un proceso particular.

Sector M. Rítmicamente muy preciso, presenta una segunda instancia de exploración de las primeras reflexiones. En el 6/4 se observa una “resíntesis” instrumental del comportamiento de los primeros ecos, propios del ambiente simulado. Significa que el comportamiento “natural” de la sala virtual es imitado por los ataques de los instrumentos.

Sector N. Conformado por dos planos. Los vientos realizan una textura contrapuntística, donde se pone de relieve la registración espacial de la altura. Los giros se entrecruzan en el espacio virtual, siguiendo diversas trayectorias, y las velocidades están en estrecha relación con la interválica de las melodías. El piano, por otra parte, ejecuta una secuencia de acordes modulante en consonancia, cuyo ritmo deviene de las relaciones temporales entre los ecos de un espacio particular, en relación con la posición de la fuente en ese ámbito. Se trata de un poderoso *zoom*, aplicado al comportamiento sonoro de una fuente en movimiento en un recinto cerrado (en este caso se trata del piano, cuyo sonido va rotando alrededor de la sala).

Sectores P y Q. Reexposición variada de diversos elementos presentados y final.

Duración promedio de *Interiores* : 25 minutos.

The musical score is presented in a standard orchestral layout with five staves: Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (Cl B), and Piano (Pf). The key signature is one flat (B-flat major or D minor), and the time signature is 4/4. The score is divided into systems, with measures 1-2, 3-8, 4, and 5. Rehearsal marks B, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are placed at the beginning of various sections. Dynamic markings include *p*, *sfzpp*, *f*, *mf*, *mp*, and *ppp*. Performance instructions such as *vib. variable ad lib.*, *frull.*, and *p subito* are used to guide the performers. The piano part includes a *tr* (trill) and *ped* (pedal) markings. The score is written in a clear, professional style with standard musical notation.

Musical score for measures 11-12. The score includes parts for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet/Bassoon (Cl B), and Piano (PI). The Flute, Oboe, and Saxophone parts feature trills (tr) and dynamic markings such as *f*, *sfz p*, *mp*, *p*, and *pp*. The Piano part includes a *ped* (pedal) marking and dynamic markings like *mf*, *f*, and *ppp*. A section marker 'C' is located above the Flute staff, and a rehearsal mark '11' is at the bottom right.

Musical score for measures 12-13. The Flute, Oboe, and Clarinet/Bassoon parts feature trills (tr) and dynamic markings such as *ppp*, *mf*, and *f*. The Piano part includes a *ped* marking and dynamic markings like *f* and *ppp*. A section marker 'Tutti' is located above the Flute staff, and a rehearsal mark '12' is at the bottom right.

Musical score for measures 13-14. The Flute part is marked *pp* and includes the instruction 'Mecánico, sin acentuaciones' (Mechanical, without accents) and a *v0* marking. The Oboe part includes a *v1* marking. The Piano part includes a *ped* marking and dynamic markings like *p*, *mf*, *sfz p*, and *pp*. Three numbered rehearsal marks (1, 2, 3) are placed above the Piano staff. A section marker 'D' is located above the Flute staff, and a rehearsal mark '13' is at the bottom left.

The image displays a musical score for five instruments: Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (Cl B), and Piano/Forte (Pf). The score is organized into three systems, each with five staves. Vertical dashed lines with upward-pointing triangles are placed above the staves to indicate specific points in time. The Flute, Oboe, and Saxophone parts feature melodic lines with various dynamics and articulations. The Clarinet Bass part includes melodic lines with dynamics such as *pp*, *p*, *f*, *mp*, and *mf*. The Piano/Forte part is the most complex, featuring a dense texture of chords and melodic fragments, with dynamics ranging from *pp* to *f*. It includes numerous performance markings such as *ped*, *sfz*, *tr*, *rall.*, and *3*. Circled numbers (1, 2, 3, 4, 5) and their primed versions (1', 2', 3', 4', 5') are placed above the Piano/Forte staff to denote specific measures or sections. The overall layout is clean and professional, typical of a printed musical score.

First system of the musical score. It includes staves for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet/Bassoon (Cl B), and Piano (Pf). The Flute part features a trill (tr) and a vibrato (v0) section. The Piano part is highly detailed with dynamic markings (f, p, mf, f, pp) and multiple pedal (ped) instructions. It also includes a trill (tr) and a circled measure with a double prime (3'').

Second system of the musical score. The Oboe part has a vibrato (v1) section. The Piano part continues with dynamic markings (mf, f, mp, sfz, p) and includes a trill (tr) and a circled measure with a double prime (5'').

Third system of the musical score. The Flute part has a vibrato (v-1) section. The Piano part includes a trill (tr) and a circled measure with a double prime (2'').

Musical score for measures 12 and 13. The score includes parts for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (ClB), and Piano/Forte (Pf). The Flute part has a dynamic marking of *mp* and a *v0* marking. The Oboe part has a *cresc.* marking. The Saxophone part has a *v-1* marking. The Clarinet Bass part has a *cresc.* marking. The Piano/Forte part has dynamic markings of *ppp*, *f*, *p*, *f*, *mp*, *f*, *p*, *mf*, and *f*. There are *ped* markings under the piano part.

Musical score for measures 14 and 15. The score includes parts for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (ClB), and Piano/Forte (Pf). The Flute part has dynamic markings of *mf*, *f*, *ppp*, *mf*, and *ppp < mp >*. The Oboe part has dynamic markings of *mf*, *f*, *ppp*, *mf*, and *ppp < mp >*. The Saxophone part has dynamic markings of *mf*, *f*, *ppp*, *mf*, and *ppp < mp >*. The Clarinet Bass part has dynamic markings of *f*, *ppp*, *mf*, and *ppp < mp >*. The Piano/Forte part has dynamic markings of *mf*, *ff*, *pp*, *mp*, and *p*. There are *ped* markings under the piano part. A box labeled 'E' contains a 4-measure rest with a tempo marking of $\text{♩} = 52$. Measure 15 has a 2-measure rest.

Musical score for measure 16. The score includes parts for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (ClB), and Piano/Forte (Pf). The Flute part has dynamic markings of *ppp < mp >*, *p*, *p*, *p*, *p*, *pp < p*, *ppp < p >*, *pp < mp >*, *pp*, and *pp < mp >*. The Oboe part has dynamic markings of *ppp < mp >*, *pp*, *pp*, *p*, *pp < p*, *pp < p >*, *pp < mp >*, and *mp*. The Saxophone part has dynamic markings of *ppp < mp >*, *mp*, *pp < p*, *pp < p*, *ppp < mp >*, *pp*, *pp < p*, and *pp < mp >*. The Clarinet Bass part has dynamic markings of *ppp < mp >*, *mp*, *p*, *mp*, *pp*, *mp*, *mp*, *ppp < mp >*, *pp*, *mf*, and *pp < mp >*. The Piano/Forte part has dynamic markings of *mf* and *p*. There is a *ped* marking under the piano part. A *rall.* marking is present above the Flute part.

F $\text{♩} = 60$
Ad libitum, espressivo

Fl
Ob
Sax
Cl B
Pn

G $\text{♩} = 52$ **H** $\text{♩} = 60$
Ad libitum, espressivo

Ob
Sax
Cl B
Pn

I $\text{♩} = 52$ **J** $\text{♩} = 60$
Ad libitum, espressivo

Fl
Ob
Sax
Cl B
Pn

Musical score for measures 21 and 22. The score includes parts for Saxophone (Sax), Flute (Fl), Oboe (Ob), Clarinet Bass (CLB), and Piano (Pf). Measure 21 features a Saxophone line with dynamics *mf*, *mf*, *f*, *pp*, *f*, *p*, *f*, *pp*, *f*, *p*. A diamond marker 's' is placed under the first *mf* dynamic. Measure 22 features a Saxophone line with dynamics *pp*, *f*, *mp*, *p*, *mp*, *pp*, *p*, *mf*, *f*, *pp*, *f*, *p*. A diamond marker 'k' is placed under the final *p* dynamic. The Flute part has a tempo marking of $\text{♩} = 40$ and dynamic *pp*. The Oboe part has dynamic *pp*. The Clarinet Bass part has a tempo marking of $\text{♩} \sim 60$ and dynamic *pp*. The Piano part has a tempo marking of $\text{♩} = 40$ and dynamic *p*. A diamond marker 'h' is placed under the final *f* dynamic. Performance instructions include *Ad libitum, espressivo* and *Rubato*. A diamond marker '22' is placed at the bottom center.

Musical score for measures 22 and 23. The score includes parts for Clarinet Bass (CLB). Measure 22 features a CLB line with dynamics *p*, *mp*, *p*, *mp*, *p*, *mf*, *p*. A diamond marker 'g' is placed under the first *p* dynamic. Measure 23 features a CLB line with dynamics *mp*, *p*, *mf*, *p*, *f*. A diamond marker 'f' is placed under the final *f* dynamic. Performance instructions include *Rubato*, *poco rall.*, *Más lento*, and *irregular, poco accel.*. A diamond marker '23' is placed at the bottom center.

Musical score for measures 23 and 24. The score includes parts for Flute (Fl), Oboe (Ob), Saxophone (Sax), and Clarinet Bass (CLB). Measure 23 features a Flute line with a tempo marking of $\text{♩} = 52$ and dynamic *ppp*. A diamond marker 'M' is placed above the tempo marking. The Oboe part has dynamics *ppp*, *mp*, *p*, *mp*, *p*, *p*. The Saxophone part has dynamics *ppp*, *mp*, *mp*, *mp*, *p*, *pp*, *p*. The Clarinet Bass part has dynamics *mp*, *p*, *f*, *mp*, *p*, *mp*, *f*. A diamond marker '23' is placed at the bottom center. Measure 24 features a Flute line with a tempo marking of $\text{♩} = 60$ and dynamic *mp*. The Oboe part has dynamics *mp*, *pp*, *mp*, *pp*, *p*, *pp*, *mf*, *pp*. The Saxophone part has dynamics *mp*, *pp*, *f*, *pp*, *mp*, *pp*. The Clarinet Bass part has dynamics *mp*, *mp*, *f*, *mf*, *pp*, *f*. A diamond marker '24' is placed at the bottom center.

First system of the musical score. It includes staves for Flute (Fl), Oboe (Ob), Saxophone (Sax), Clarinet Bass (Cl B), and Piano (Pf). The Flute part features a melodic line with dynamics *mp*, *p*, *mp*, *p*, *mf*, and *pp*. The Oboe part has dynamics *p* and *mp*. The Saxophone part has dynamics *p* and *mp*. The Clarinet Bass part has dynamics *pp*, *p*, and *mf*. The Piano part includes a bass line with several *ped* (pedal) markings.

Second system of the musical score. The Flute part has dynamics *p*, *mf*, and *pp*. The Oboe part has dynamics *p* and *mf*. The Saxophone part has dynamics *p*, *mp*, and *mf*. The Clarinet Bass part has dynamics *p* and *mf*. The Piano part continues with *ped* markings.

Third system of the musical score. The Flute part has dynamics *p* and *mf*. The Oboe part has dynamics *pp* and *p*. The Saxophone part has dynamics *p* and *mp*. The Clarinet Bass part has dynamics *mf* and *p*. The Piano part continues with *ped* markings.

Flute (Fl): *mf*, *p*, *mf*, *mf*, *p*

Oboe (Ob): *p*, *mf*, *mf*, *p*

Saxophone (Sax): *p*, *mf*, *mp*, *p*

Clarinet Bass (Cl B): *p*, *mf*, *mp*, *mf*, *mp*, *p*

Piano (Pf): *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*

Flute (Fl): *p*

Oboe (Ob): *p*, *mp*

Saxophone (Sax): *p*, *mp*

Clarinet Bass (Cl B): *mp*, *pp*, *p*

Piano (Pf): *p*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*

Flute (Fl): *p*, *mp*, *p*, *mf*

Oboe (Ob): *p*, *mp*, *p*, *mf*

Saxophone (Sax): *p*, *mp*, *p*, *mf*

Clarinet Bass (Cl B): *p*, *mf*, *mp*, *mf*, *p*

Piano (Pf): *mp*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*, *ped*

Fl
Ob
Sax
Cl B
Pf

Fl
Ob
Sax
Cl B
Pf

Fl
Ob
Sax
Cl B
Pf

Pf

ped ped

Pf

ped ped

Pf

ped ped

5 $\text{♩} = 52$ **P** poco a poco accel.

Fl $\text{♩} = 72$ 2 *tr*

Ob *tr*

Sax *tr*

ClB *tr*

Pf *Rubato* *pp* *mf* *f* *pp* *ped*

27

3 $\text{♩} = 60$ 2

Fl *p* *f* *pp*

Ob *p* *f* *pp* *p < mp*

Sax *pp* *mf* *p* *f* *sempre pp*

ClB *pp* *mf* *f* *sempre pp* *v0*

Pf *p* *ped* *mp* *ped* *f* *ped* *Rubato* *pp* *ped* *f*

28

52

Flute I (Fl) and Oboe (Ob) parts feature trills and dynamic markings *p*, *mp*, and *p*. Saxophone (Sax) and Clarinet Bass (Cl B) parts play a rhythmic pattern with *vi* and *vo* markings. The Piano (Pf) part includes chords and a *ped* marking.

3 *accel.*

Measures 29-31. Flute I and Oboe parts have trills and dynamics *p* and *f*. Saxophone and Clarinet Bass parts have *vi* and *vo* markings. Piano part includes *pp*, *f*, and *sfz/pp* markings, with a *ped* marking and a circled measure number 29.

3 Q

Measures 30-31. Flute I and Oboe parts feature trills and dynamics *pp* and *ppp*. Saxophone and Clarinet Bass parts have *pp* and *ppp* markings. Piano part includes *p*, *mp*, *p*, and *espressivo* markings, with *ped* markings and circled measure numbers 30 and 31.

Piano (Pf) part, measures 30-31, featuring chords and *ped* markings.

3.6 Conclusiones

Partimos de las teorías psicoacústicas a fin de comprender los fenómenos ligados a la percepción auditiva, y analizamos distintos modos de simulación de fuentes y ambientes virtuales, con el propósito de estudiar sus principios, aciertos y deficiencias, y llevarlos al ámbito del procesamiento en tiempo real. Toda esta información sirvió de base para el diseño y desarrollo de un entorno de composición, que asiste en la creación de obras musicales.

La utilización de técnicas binaurales, extendidas a la cuadrafonía, dio por resultado la posibilidad de generar imágenes sonoras en tres dimensiones, mediante un número muy limitado de canales. Si bien la ubicación del oyente en el espacio de audición condiciona el nivel de precisión en la localización de los eventos auditivos, el empleo de cuatro canales dispuestos al frente y atrás, permitió extender el área útil destinada al auditorio, y resolver ambigüedades en la percepción, propias de la reproducción transaural. El sistema, utilizado en la composición de una obra musical, presentó resultados satisfactorios, que deberán ser confirmados luego, a través de métodos formales.

El dispositivo de localización espacial, y otras herramientas de asistencia en la composición, fueron integrados a un mismo entorno, con posibilidades de expansión sólo limitadas por la capacidad de procesamiento de la plataforma elegida. Tal integración brindó la posibilidad de realizar la interconexión de todas ellas de forma inmediata, y a la vez, permitió la puesta en práctica de las ideas compositivas en estrecho vínculo con el tratamiento espacial del sonido. De este modo, la utilización del espacio en música encontró su justificación inmediata, desde un punto de vista estrictamente musical.

El tipo de entorno de desarrollo elegido permitió modificaciones rápidas y configuraciones variadas de las aplicaciones. La expansión del sistema podrá ser realizada por cualquier usuario, ya que la documentación de los objetos y sus interconexiones es precisa, y responde a un formato estándar, provisto por el mismo compilador acústico.

El funcionamiento en tiempo real del sistema ofrece múltiples usos. Si bien fue concebido originalmente para la composición musical con medios mixtos, desarrollos multimediáticos, instalaciones y otras aplicaciones artísticas, presentan campos alternativos de utilización.

Las ideas compositivas aquí planteadas, y plasmadas en la composición de la obra, generan puntos de partida interesantes para los próximos desarrollos. La integración de los aspectos temporales y espaciales de la música, la exploración de su espacio interno en relación al espacio externo, la creación de programas que hagan posible una música interactiva basada en las decisiones de un oyente, proponen desafíos nuevos, y deben ser profundizadas en el futuro.

Referencias bibliográficas

Referencias bibliográficas

Algazi, V., Avendano, C. y Duda, R. (2001). "Elevation localization and head-related transfer function análisis at low frequencies". J.A.S.A. 109(3), 1110-1122.

Algazi, V., Duda, R. y otros (2002). "Approximating the head-related transfer function using simple geometric models of the head and torso". J.A.S.A. 112(5), 2053-2064.

Allen, J. B. y Berkley, D. A. (1979). "Image method for efficiently simulating small room acoustics". J.A.S.A. 65(4).

Batra, R., Kuwada, S. y Stanford, T. (1989). "Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit". J. Neurophysiol. 61, 257-268.

Batteau, D. W. (1967). "The role of the pinna in human localization". Proc. Roy. Soc. London. B-168, 158-180.

Batteau, D. W. (1968). "Listening with the naked ear". En "The neuropsychology of spatially oriented behavior". S. J. Freedman (Ed). Dorsey Press, Homewood, IL. pp. 109-133.

Bauer, B. (1961). "Stereophonic earphones and binaural loudspeakers". J.A.E.S. 9, 148-151.

Begault, D. R. (1987). "Control of auditory distance". Doctoral dissertation. Department of Music. UCSD.

Begault, D. R. (1991). "Perceptual effects of synthetic reverberation on 3D audio systems". J.A.E.S. 40, 895-904.

Berkhout, A. J. (1988). "A holographic approach to acoustic control". J.A.E.S. 88, 977-995.

Berkhout, A. J., Vries, D. y Vogel, P. (1993). "Acoustic control by Wave Field Synthesis". J.A.E.S. 93, 2764-2778.

Blauert, J. (1997) "Spatial Hearing. The psychophysics of human sound localization". The MIT Press. Cambridge, Massachusetts.

Boerger, G. (1965). "On the persistence of the directional sensation in hearing". Proceedings, 5th International Congress on Acoustics, Liège, B 27.

Butler, R. A. (1987). "An analysis of the monaural displacement of sound in space". Percept. Psychophys.. 41, 1-7

- Butler, R. A. (1997). "Spatial referents of stimulus frequencies: their role in sound localization". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. pp 77-97.
- Butler, R., Humanski, R. y Musicant, A. (1990). "Binaural and monaural localization of sound in two dimensional space". *Perception*. 19, 241-256.
- Cetta, P. (1996). "Estudio y desarrollo de una interfaz gráfica a aplicar en programas de síntesis digital del sonido, destinada a composición y docencia". Fondo Nacional de las Artes. Becas de Investigación 1996.
- Cetta, P. (2003a) "Modelos de localización espacial del sonido y su implementación en tiempo real". Altura-Timbre-Espacio. Instituto de Investigación Musicológica "Carlos Vega". FACM-UCA.
- Cetta, P. (2003b) "Principios de estructuración de la altura empleando conjuntos de grados cromáticos". Altura-Timbre-Espacio. Instituto de Investigación Musicológica "Carlos Vega". FACM-UCA.
- Chandler, D. y Grantham, D. (1992). "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth and velocity". *J.A.S.A.* 91, 1624-1636.
- Chernyak, R. y Dubrovsky, N. (1968). "Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise". *Proceedings, 6th International Congress on Acoustics, Tokyo*, A 3-12.
- Chowning, J. (1971). "The simulation of moving sound sources". *J.A.E.S.* 19, 2-6.
- Chowning, J. (1990). "Music from machines : Perceptual fusion and auditory perspective". Report STAN M-64. CCRMA. Stanford University.
- Cooper H. y Bauck, J. (1989). "Prospects for transaural recording". *J.A.E.S.* Vol. 37. pp. 3-19.
- Cremer, L. (1948). "Die wissenschaftlichen Grundlagen der Raumakustik [The scientific foundations of architectural acoustics]". Vol 1. S. Hirzel Verlag, Stuttgart.
- Dadoun, N., Kirkpatrick, D. y Walsh, J. (1985). "The geometry of beam tracing". *Proc. of the Symposium on Computational Geometry 1985*. pp 55-71.
- Damaske, P. (1967/8). "Subjektive Untersuchungen von Scalfeldern [Subjective investigations of sound fields]. *Acustica* 19, 198-213.

- Damaske, P. y Wagener, B. (1969) . "Richtungshörversuche über einen nachgebildeten Kopf [Investigations on directional hearing using a dummy head]. *Acustica* 21, 30-35.
- David, E., Guttman, N. y van Bergeijk, W. (1958). "On the mechanism of binaural fusion". *J.A.S.A* 30, 801-802.
- Davis, D. y Davis, C. (1980). "The LEDE concept for the control of acoustic and psychoacoustic parameters in recording control rooms". *J.A.E.S.* 28, 585-595.
- Di Liscia, O. P. (1999) "RTSPA1 a program and some remarks on real time spatialization of sound". *Registros históricos del VI Simposio Brasileño de Música y Computadoras, Río de Janeiro, Brasil.*
- Duda, Richard (1997). "Elevation Dependence of the Interaural Transfer Function". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. 1997.
- Elfner, L. y Perrot, D. (1967). "Effect of prolonged exposure to a binaural intensity mismatch on the locus of a dichotically produced tonal image" *J.A.S.A.* 39, 716-719.
- Feddersen, W., Sandel, T., Teas, D. y Jeffress, L. (1957). "Localization of high frequency tones". *J.A.S.A.* 29, 988-991.
- Fisher, H. y Freedman, S. (1968). "The role of the pinna in auditory localization". *J. Auditory Res.* 8, 15-26.
- Gardner, M. B. (1973). "Problem of localization in the median plane: effect of pinna cavity occlusion". *J.A.S.A.* 53, 400-408.
- Gardner, M. B. (1969). "Distance estimation of 0° or apparent 0° oriented speech in anechoic space". *J.A.S.A.* 45, 47-53.
- Gardner, W. G. (1995). "Efficient convolution without input-output delay". *J.A.E.S.* 43, 127-136.
- Gardner, W. G., Martin, K. (1994). "HRTF Measurements of a KEMAR Dummy-Head Microphone". *MIT Media Lab Perceptual Computing - Technical Report #280*
- Gerzon, M. A. (1976). "Unitary (energy preserving) multichannel networks with feedback". *Electronic Letters.* 12 (11).
- Gerzon, M. A. (1985). "Ambisonics in multichannel broadcasting and video". *J.A.E.S.* 33(11), 859-871.

- Goldberg, J. y Brown, P. (1969) "Response of binaural neurons of dog superior olivary complex dichotic tonal stimuli: some physiological mechanisms of sound localization". *J. Neurophysiol.* 32, 613-636.
- Green, D. M. (1971). "Temporal auditory acuity". *Psych. Rev.* 78, 540-551.
- Green, D. M. (1976). "An introduction to hearing". Lawrence Erlbaum Associates, Hilldale, NJ.
- Griesinger, D. (1989). "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction". *J.A.E.S.*, 37(1/2), 20-29.
- Haas, H. (1951). "The influence of a single echo in the audibility of speech". Traducción del original alemán, publicada en 1972 en *J.A.E.S.* 20, 146-159.
- Hartley, R. y Fry, T. (1921). "The binaural location of pure tones". *Phys. Rev.* 18, 431-442.
- Hartmann, W. M. (1997). "Listening in a room and the precedence effect". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. pp 191-210.
- Hartmann, W. y Wittenberg, A. (1996). "On the externalization of sound images". *J.A.E.S.* 99, 3678-3688.
- Henry, J. (1849). Citado en Blauert (1997).
- Jérôme, D., y otros (2003). "Further Investigations of high order Ambisonics and Wavefield Synthesis for Holophonic sound imaging". *J.A.E.S.* ntion Paper 5788. Presented at 114th Convention, Amsterdam.
- Jot, J. M. (1991). "Digital delay networks for designing artificial reverberators". *Proc. 90th Conv. A.E.S.*
- Jot, J. M. (1996). "Synthesizing three-dimensional sound scenes in audio or multimedia production and interactive human-computer interfaces". 5th International Conference "Interface to real and virtual worlds" (Montpellier, France).
- Jot, J. M. (1997). "Efficient models for reverberation and distance rendering in computer music and virtual audio reality". *Proc. ICMC 1997.*
- Julien, P., Kahle, E., Winsberg, S. y Warusfel, O. (1992). "Some results on the objective and perceptual characterization of room acoustical quality in both laboratory and real environments". *Proc. Institute of Acoustics.* XVI (2).

- Kendall, G. (1995). "The decorrelation of audio signals and its impact on spatial imagery". *Computer Music Journal* 19 (4), 71-87.
- Kendall, G. y Martens, W. (1984). "Simulating the cues of spatial hearing in natural environments". *Proc. of the International Computer Music Conference. IRCAM. Paris.* pp. 111-125.
- Krockstadt, U. R. (1968). "Calculating the acoustical room response by the use of a ray tracing technique". *J. Sound and Vibrations.* 8(18).
- Kuhn, G. F. (1987). "Physical acoustics and measurements pertaining to directional hearing". En "Directional Hearing", editado por Yost y Gourevitch. Springer Verlag. New York.
- Kyriakakis, C. (1998). "Fundamental and technological limitations of immersive audio systems". *Proc. of the I.E.E.E.*, 86 (5), 941-951.
- Malham, D. G. y Myatt, A. (1995). "3D sound spatialization using Ambisonic techniques". *Computer Music Journal* 19(4), 58-70.
- Malham, D. G. (1998). "Spatial hearing mechanisms and sound reproduction". <http://www.york.ac.uk/inst/mustech/welcome.htm>.
- Malham, D.G. (1999) "Higher order Ambisonics systems for the spatialization of sound". *Proceedings ICMC 1999. Beijing.*
- Malham, D.G. (2002) "Second and Third Order Ambisonics - the Furse-Malham Set". http://www2.york.ac.uk/inst/mustech/3d_audio/secondor.html
- Middlebrooks, J. C. y Green, D. M. (1990). "Directional dependence of interaural envelope delays". *J.A.S.A* 87, 2149-2162.
- Middlebrooks, J. C. (1997). "Spectral shape cues for sound localization". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. pp 77-97.
- Mills, A. W. (1958). "On the minimum audible angle". *J.A.E.S* 30, 237-246.
- Möller, H. (1989). "Cancellation of crosstalk in artificial head recordings reproduced through loudspeakers". *J.A.E.S.*, vol. 37, pp. 31-34.
- Moore, B., Oldfield, S., y Dooley, G. (1989). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz". *J.A.S.A.* 85, 820-835.

- Moore, F. R. (1978a). "An introduction to the mathematics of digital signal processing, part 1". *Computer Music Journal* vol. 2 no. 1, 38-47.
- Moore, F. R. (1978b). "An introduction to the mathematics of digital signal processing, part 2". *Computer Music Journal* vol. 2 no. 2, 38-60.
- Moore, F. R. (1983). "A general model for spatial processing of sound". *Computer Music Journal* 7(3), 6-15.
- Moore, F. R. (1990). "Elements of computer music". Prentice-Hall. Englewood Cliffs, New Jersey.
- Moore, F. R. (1991). "Spatialization of sounds over loudspeakers". En M. Mathews y J. Pierce (eds.) "Current directions in computer music research". MIT Press. Cambridge, Massachusetts. pp 89-104.
- Moorer, J.A. (1979). "About this reverberation business". *Computer Music Journal*. 3(2), 13-28.
- Morimoto, M. y Aokata, H. (1984). "Localization cues of sound sources in the upper hemisphere". *J. Acoust. Soc. Japan* 5, 165-173.
- Musicant, A., y Butler, R. (1984). "The influence of pinnae-based spectral cues on sound localization". *J.A.S.A.* 75, 1195-1200.
- Musicant, A. D. y Butler, R. A. (1985). "Influence of monaural spectral cues on binaural localization". *J.A.E.S.* 77, 202-208.
- Pratt, C. (1930). "The spatial character of high and low tones". *J. Exp- Psychol.* 13, 278-285.
- Puckette, M. (1991). "Combining event and signal processing in the Max graphical programming environment". *Computer Music Journal*, 15 (3).
- Pulkki, V. (1997). "Virtual sound source positioning using vector based amplitude panning". *J.A.E.S.* 45(6), 456-466.
- Pulkki, V. (1999). "Uniform spreading of amplitude panned virtual sources". *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct., 17-20*
- Pulkki, V. (2000). "Generic panning tools for MAX/MSP". <http://www.acoustics.hut.fi>.
- Rayleigh, Lord (1907). "On our perception of sound direction". *Philos. Mag.* 13, 214-232.
- Rayleigh, Lord (1877) "Acoustical observations". *Philos. Mag.* 3, 456-464.

- Rife, D. D. y Vanderkooy, J. (1989). "Transfer-Function Measurements using Maximum-Length Sequences". *J.A.E.S* 37(6), 419-444.
- Roffler, S. y Butler, R. (1968). "Localization of tonal stimuli in the vertical plane". *J.A.S.A.* 43, 1260-1266.
- Rogers, M. E. y Butler, R. A (1992). "The linkage between stimulus frequency and covert praks areas as it relates to monoaural localization". *Percept. Psychophys.* 52, 536-546.
- Rose, J., Gross, N., Geisler, C. y Hind, J. (1966). "Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source". *J. Neurophysiol.* 29, 288-314.
- Sayers, B. (1964). "Acoustic image lateralization judgements with binaural tones". *J.A.S.A.* 36, 923-926.
- Scherer (1959). Citado por Blauert (1997).
- Schroeder, M. R. (1962). "Natural sounding artificial reverberation". *J.A.E.S.* 10(3), 219-223.
- Schroeder, M. y Atal, B. (1963). "Computer simulation of sound transmission in rooms". *IEEE Conv. Rec.*, pt 7, pp. 150-155.
- Searle, C. L y otros (1975). "Binaural pinna disparities: another auditory localization cue". *J.A.E.S.* 57, 448-435.
- Shaw, E. (1966). "Ear canal pressure generated by a free sound field" *J.A.S.A.* 39, 465-470.
- Shaw, E. (1997). "Acoustical features of the human external ear". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. pp 25-47.
- Shaw, E. A. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane". *J.A.E.S.* 56, 1848-1861.
- Shaw, G. y Teranishi, R. (1968). "Sound pressure generated in an external ear replica and real human ears by a nearby sound source". *J.A.S.A.* 44, 240-249.
- Shinn-Cunningham, B. (2000). "Learning reverberation: implications for spatial auditory displays". *Internacional Conference on Auditory Displays*, Atlanta, GA.
- Slattery, W. y Middlebrooks, J.C. (1994). "Monoaural sound localization: acute versus chronic unilateral impairment". *Hear. Res* 75, 38-46.

- Smith, Julius (1985). "Introduction to Digital Filter Theory". Report STAN M-20. CCRMA. Stanford University.
- Snow, W. (1953). "Basic principles of stereophonic sound". *SMPTE Journal*, 61, 567-589.
- Stautner, J. y Puckette, M. (1982). "Designing multi-channel reverberators". *Computer Music J.* 6(1).
- Stevens, S. y Newman, E. (1936). "The localization of actual sources in sound". *Am. J. Psych.* 48, 297-306.
- Stockhausen, K. (1956). "Programme notes for the 1956 world premiere of *Gesang der Jünglinge*". CD *Stockhausen elektronische musik, 1952-1960*. Stockhausen Verlag.
- Thiele, G. (1993). "The new sound format 3/2-stereo". *Proc. 94th Conv. Audio Eng. Soc.*, preprint 3550a.
- Tobias, J. y Serlin, S. (1959). "Lateralization thresholds as a function of stimulus duration". *J.A.E.S.* 31, 1591-1594.
- Tonning, F. M. (1970). "Directional audiometry, I. Directional white noise audiometry". *Acta Otolaryngol.* 69, 388-394.
- Toole, F. y Sayers, B. (1965). "Lateralization judgements and the nature of binaural acoustic images". *J.A.S.A.* 37, 319-324.
- van Soest, J. (1929). "Directional hearing of sinusoidal sound waves". *Physica* 9, 271-282.
- Varios (1994). "Espaces". *Les cahiers de l'IRCAM. Recherche et musique* 5.
- Vercoe, B. (1993). "Csound. Software and manual". MIT. Cambridge. Massachussets.
- von Békésy, G. (1949). "The moon illusion and similar auditory phenomena". *Am J. Psychol.* 62, 540-552.
- von Békésy, G. (1971). "Auditory backward inhibition in concert halls". *Science* 171, 529-536.
- von Hornbostel, E. M. y Wertheimer, M. (1920). "Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]". *Sitzungsber. Akad. Wiss. Berlin*, 388-396.
- Wagener, W. (1971). "Räumlich Verteilungen der Hörrichtungen in synthetischen Scalfeldern [Spatial distribution of auditory directions in synthetic sound fields]". *Acustica* 25, 203-219.

Wallach, H., Newman, E. y Rosenzweig, M. (1949). "The precedence effect in sound localization". *Am. J. Psych.* 62, 315-337.

Wallach, H., Newman, E. y Rosenzweig, M. (1949). "The precedence effect in sound localization". *Am J. Psychol.* 57, 315-336.

Wendt, K. (1963). "Directional hearing with two superimposed sound fields in intensity and delay difference stereophony". Disertación. Technische Hochschule. Aachen.

Wiener, F. (1947). "On the diffraction of a progressive sound wave by the human head". *J.A.S.A.* 19, 143-146.

Wightman, F., y Kistler, D. (1992). "The dominant role of low-frequency interaural time difference in sound localization". *J.A.S.A.* 91, 1648-1661.

Wightman, F., y Kistler, D. (1997). "Factors affecting the relative salience of sound localization cues". En "Binaural and Spatial Hearing in Real and Virtual Environments". Editado por Gilkey y Anderson. Lawrence Erlbaum Associates. New Jersey. pp 1-23.

Woodworth, R. y Schlosberg, H. (1954). "Experimental Psychology". Holt. New York.

Yadegari, S., Moore, F. R., Burr, A., Castle, H. y Apel, T. (2002). "Real-Time Implementation of a General Model for Spatial Processing of Sounds". Proceedings, International Computer Music Conference, Goteborg, Sweden. San Francisco: International Computer Music Association. pp. 244-247.