



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA

**Programa para reproducción de sonido multicanal
en tiempo real y síntesis de frentes de onda (WFS).**

T E S I S

Que para obtener el título de:
MAESTRO EN INGENIERÍA ELÉCTRICA
INSTRUMENTACIÓN ACÚSTICA

P R E S E N T A
HUGO EZEQUIEL GUTIÉRREZ HERNÁNDEZ

TUTOR

DR. FELIPE ORDUÑA BUSTAMANTE

2010





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Resumen

Esta tesis trata sobre el desarrollo de una serie de programas de cómputo enfocados a la reproducción de sonido multicanal en tiempo real y a la implementación de la técnica de Síntesis de Frentes de Onda (WFS, sus siglas en inglés). Estos programas utilizan como material de reproducción, señales grabadas o en vivo, de uno o varios canales, para enviarlas a un arreglo lineal de ocho altavoces, después de aplicarles un procesamiento digital de señales. Por otra parte se ha investigado la posibilidad de generar frentes de onda de manera sintética, bien para reforzar u modificar algunos ya existentes o para la generación de fuentes virtuales de características diversas. Mediante el procesamiento desarrollado se logran varios efectos: circular señales secuencialmente en los ocho altavoces del arreglo, y efectos de lateralización de la señal reproducida, queriendo decir con esto que la señal parece emitirse con un cierto ángulo respecto a la normal del arreglo, con distintos grados de lateralidad que se controlan desde el software; estos son los más importantes, por sus posibles aplicaciones. Al utilizar señales estereofónicas, y mediante los efectos de lateralización, se permite expandir el área de escucha, comparada con un sistema convencional de dos canales, de modo que se tiene un sistema con menor dependencia de la posición del oyente. Como herramientas de programación, los programas desarrollados se valen del uso de una plataforma de software libre (principalmente, las bibliotecas `fftw3` y `portaudio`) y como elemento de procesamiento, utilizan una interfaz de audio apropiada para el manejo de ocho altavoces. Esto le da a esta implementación la ventaja de ser independiente de los sistemas operativos de cómputo. Se llevaron a cabo pruebas objetivas (mediciones con arreglos de micrófonos, análisis espectrales y espaciales) y subjetivas para su evaluación. Éstas últimas ofrecen resultados alentadores y abren la posibilidad de utilizar dichos programas en aplicaciones donde la calidad de la reproducción estéreo mejora, con efectos perceptibles distribuidos dentro de una más amplia zona de escucha que en los sistemas estereofónicos convencionales, en beneficio de un mayor número de oyentes.

Reconocimientos

Lista de figuras

- Figura 2.1 Diferencias interaurales de tiempo y amplitud en es sistema biaural de audición humano. 7
- Figura 2.2 Colaboración del director Leopold Stokowski con Harvey Fletcher en la reproducción de audio. 13
- Figura 2.3 Ingeniero Alan Blumlein. Contribuciones al sistema estereofónico. 14
- Figura 2.4 Implementación del sistema estereofónico ideal. 20
- Figura 2.5 Implementación del sistema estereofónico de tres canales. 21
- Figura 2.6 Efecto del cambio de posición en la imagen fantasma en el sistema estereofónico. 23
- Figura 2.7 Arreglo lineal de altavoces para Síntesis de Frentes de Onda WFS. 27
- Figura 2.8 Ilustración del objetivo fundamental de la técnica WFS. 28
- Figura 2.9 Geometría para la integral de Kirchhoff-Helmholtz. 29
- Figura 2.10 Geometría para la integral de Kirchhoff-Helmholtz para superficie que degenera en un plano. 30

• Figura 2.11 Diagrama de implementación práctica de WFS.	31
• Figura 2.12 Configuración de altavoces para WFS.	33
• Figura 3.1 Geometría utilizada en el análisis del arreglo lineal de fuentes acústicas	38
• Figura 3.2 Representación en curvas de nivel para el campo de presión sonora instantánea para un arreglo de 8 altavoces y direccionalidad de 0°.	43
• Figura 3.3 Representación en curvas de nivel para el campo de presión sonora instantánea para un arreglo de 8 altavoces y direccionalidad de 30°.	44
• Figura 3.4 Representación en curvas de nivel para el campo de presión sonora instantánea para un arreglo de 8 altavoces y direccionalidad de 60°.	45
• Figura 3.5 Representación en curvas de nivel para el campo de presión sonora instantánea para un arreglo de 8 altavoces y direccionalidad de 90°.	46
• Figura 3.6 Diagrama general del sistema implementado.	47
• Figura 3.7 Esquema del programa 'circulación.c'.	48
• Figura 3.8 Esquema del programa 'retardosizq.c'.	49
• Figura 3.9 Esquema del programa 'retardosder.c'.	50
• Figura 3.10 Esquema del programa 'retizqder.c'.	51
• Figura 3.11 Esquema del programa 'estereo.c'.	52
• Figura 3.12 Esquema del programa 'estereom.c'.	53
• Figura 4.1 Arreglo lineal de altavoces y de micrófonos para la evaluación experimental del	54

sistema propuesto.

- Figura 4.2 Respuestas a impulso entre la señal de entrada de audio monofónica y la captada por los micrófonos en condiciones de campo libre. 56
- Figura 4.3 Magnitud de las respuestas en frecuencia de la señal de entrada de audio monofónica y la captada por los micrófonos en condiciones de campo libre. 57
- Figura 4.4 Fase de las respuestas en frecuencia entre la señal de entrada de audio monofónica y la captada por los micrófonos en condiciones de campo libre. 58
- Figura 4.5 Retraso de grupo de las respuestas en frecuencia entre la señal de entrada de audio monofónica y la captada por los micrófonos en condiciones de campo libre. 59
- Tabla 2.1 Evaluación de las técnicas de reproducción estereofónica, transaural y ambisónica para localización de fuentes sonoras. 25
- Tabla 3.1 Relación entre el tiempo de retardo aplicado a los altavoces del arreglo lineal y el ángulo de direccionalidad del arreglo. 41

Índice

Resumen	I
Reconocimientos	Ii
Lista de figuras	Ii
Indice	V
1. Introducción y alcance	1
1.1. Preliminares.	1
1.2. Potencial y limitaciones del presente trabajo.	3
1.3. Contenido de la tesis.	4
2. Antecedentes	6
2.1. Sistema de audición biaural.	6
2.1.1. Localización angular	6
2.1.2. Localización de profundidad	9
2.2. Sistemas de reproducción de sonido.	9
2.2.1. Monoaural	11
2.2.2. Biaural	11
2.2.3. Estereofónico	14
2.2.4. Cuadrafónico	16
2.2.5. Sistemas envolventes (surround)	17
2.3. Sistema Estereofónico Convencional	19
2.3.1. Condiciones de operación	20
2.3.2. Percepción angular	21
2.3.3. Efecto de la posición del observador	22
2.3.4. Número de canales	24
2.4. Límites y comparación de los sistemas de reproducción sonora.	24
2.5. Síntesis de Frentes de Onda: WFS, Wave-Field Synthesis.	26
2.5.1. Integral de Kirchhoff-Helmholtz.	28
2.5.1.1 Primer esquema de la integral de Rayleigh	31
2.5.1.2 Segundo esquema de la integral de Rayleigh	32

2.5.2	Obtención de las señales WFS.	33
2.5.3	Potencial y limitaciones.	35
2.5.3.1	Ventajas	35
2.5.3.2	Limitaciones	36
2.6	Resumen	37
3.	Implementación del sistema	38
3.1.	Arreglo lineal de altavoces.	38
3.2.	Simulaciones.	41
3.3.	Software e interfaz de audio.	47
3.4.	Programas implementados para la generación de señales acústicas.	47
3.4.1.	circulación.c	48
2.5.2	retardosizq.c	49
2.5.3	retardosder.c	49
3.4.4.	retizqder.c	50
3.4.5.	estereo.c	51
3.4.6.	estereom.c	52
3.5.	Resumen	53
4.	Resultados	54
4.1.	Pruebas objetivas y mediciones.	54
5.	Conclusiones	60
5.1.	Perspectivas	61
6.	Bibliografía	62
7.	Glosario	67
8.	Apéndices	69
8.1.	Código de los programas implementados.	69

1.Introducción y alcance

1.1. Preliminares

El objetivo de un sistema de reproducción de audio de alta calidad debe ser transmitir fielmente las características espaciales y temporales de un evento acústico grabado, de modo que los oyentes tengan las sensaciones y perspectivas que tendrían al haber estado en dicho evento. En el caso de sistemas para reforzamiento de eventos en vivo, la reproducción no debe alterar las percepciones de los oyentes, sino que dicho sistema únicamente deberá mejorar las características de potencia y permitir que la audición pueda ser efectuada por un número mayor de oyentes.

Con esta intención, y obedeciendo a las limitaciones tecnológicas de cada época particular, en el desarrollo de la ingeniería de audio se han implementado sistemas de reproducción de sonido con diferentes enfoques. En el transcurso de dichos desarrollos, grandes avances se han hecho tanto en el objetivo de una reproducción realista de eventos acústicos, como en el conocimiento del sistema de audición humano y diferentes efectos psicoacústicos, así como en el desarrollo tecnológico de sistemas de procesamiento, altavoces y micrófonos.

Entre las primeras implementaciones de sistemas de reproducción, se contó con el sistema monofónico, que mediante un solo canal de reproducción, intentaba dar una perspectiva de eventos sonoros grabados. Dicho sistema, aunque comercialmente duró mucho tiempo en uso, desde finales del siglo XIX y hasta casi mediados del siglo XX, tenía muchas deficiencias en cuanto a dar una verdadera perspectiva espacial, toda vez que el sonido provenía de una única fuente. Sin embargo, permitió el desarrollo de técnicas de grabación, incluyendo el uso de arreglos de micrófonos [45].

El sistema binaural de reproducción, históricamente casi tan antiguo como el monofónico, es otra aproximación a la reproducción realista de eventos acústicos. Dicho sistema aspira a presentar en los oídos del oyente las señales que debería presentarse habiendo estado realmente en la escena. Para ello, provee durante la grabación, las alteraciones que sufren dichas señales en el torso, barbilla y oído externo del oyente, valiéndose de maniqués especialmente diseñados o colocando micrófonos en los oídos de un oyente particular. El efecto logrado con este sistema es superior al monofónico y da una perspectiva espacial real. Sin embargo, presenta problemas debido a que las alteraciones sufridas por las señales en su camino a los oídos son únicas para cada individuo; el uso de

audífonos es incómodo y presenta algunos problemas de localización tanto hacia el frente como hacia detrás y tiende a presentar la sensación de sonido “dentro de la cabeza”.

El sistema estereofónico ha sido la aproximación comercialmente más factible y se ha mantenido durante un buen número de años en uso, desde mediados del siglo XX y aún hasta nuestros días. Aunque sus orígenes se remontan a las primeras décadas del siglo XX, la tecnología para implementarlo no estuvo disponible comercialmente sino hasta a mediados de dicho siglo. Mediante el uso de varios canales de grabación y varios de reproducción, el sistema estereofónico produce una perspectiva espacial y provee una buena localización de fuentes acústicas a través de la creación perceptual de imágenes o fuentes virtuales (fantasmas). Esto sin embargo, es sólo dentro de un área de escucha (sweet spot) que es, en general, reducida y fuera de ella la imagen acústica se degrada. Aún cuando la localización se preserva, esto no es siempre en la misma posición, haciendo que el sistema tenga un efecto antinatural respecto a la audición en vivo. Su uso es tan extendido, que se utiliza en radiodifusión de señales FM y de televisión.

A partir del sistema estereofónico han surgido otros que intentan resolver las deficiencias de las que adolece dicho sistema, utilizando la idea básica de aquél. Así pues, ha habido enfoques como la cuadrafonía, que utilizaba 4 canales de reproducción, fuera a partir de la mezcla estereofónica normal, o mediante una matriz de codificación-decodificación de 4-2-4 canales. Dicho sistema lograba tener un carácter envolvente y buena localización frontal y trasera, pero deficiente a los lados del oyente.

Otros sistemas que igualmente han logrado tener éxito comercial en los últimos años y ganado lugar en los estándares actuales, sobre todo en el ámbito cinematográfico, son los de carácter envolvente o surround. Los más populares de ellos son el 5.1 y el DTS, el primero de ellos ha sido adoptado como estándar para el sistema de televisión de alta definición (HDTV) en E. U. También existen otros formatos que utilizan más canales, pero cuyo uso no está tan extendido para consumo doméstico (surround 7.1, 9.1, 10.2, etc.). Mediante el uso de más canales, estos sistemas logran aumentar el carácter envolvente de la reproducción y al utilizar al menos 3 canales frontales, hacen que la localización de fuentes acústicas sea más precisa. Asimismo, agregan canales traseros (surround) que proporcionan efectos de ambiente y aumentan el carácter envolvente del sonido. Estos sistemas, a pesar de aumentar el carácter envolvente, presentan problemas de degradación al alejarse de la zona óptima de audición (sweet spot) y por no ser capaces de reproducir una perspectiva espacial real.

Además de los sistemas envolventes descritos anteriormente, existen otros sistemas más ambiciosos que pretenden dar mayor realismo a las señales reproducidas. El más importante de estos sistemas es el de Síntesis de Frentes de Onda o WFS, el cual, mediante arreglos extensos

de altavoces, pretende reconstruir un frente de onda basándose en el principio de Huygens de propagación ondulatoria. Mediante el procesamiento aplicado a las señales enviadas a cada altavoz del arreglo, se logra sintetizar con mucha precisión, campos acústicos complejos y eliminar las limitaciones de la zona de audición óptima. El realismo obtenido de este modo es muy superior a los de los sistemas convencionales, de modo que todos los oyentes dentro de la zona de audición, tienen la misma perspectiva espacial. Sin embargo, para lograr esto, el sistema requiere de un gran número de altavoces y un procesamiento costoso, que lo hace impráctico para el consumo doméstico [13] [20].

Hay pues, a la fecha, diferentes maneras de abordar el problema de la reproducción de campos sonoros. Dependiendo de la aplicación y complejidad del sistema los resultados varían de pobres a bastante satisfactorios. Sin embargo, el objetivo de representar una escena sonora con el realismo de presenciarla en vivo, no se ha logrado aún de modo satisfactorio, sea que esto se deba al costo del sistema, o a la calidad obtenida. Debe encontrarse una solución que presente resultados satisfactorios en términos de percepción y a la vez viable desde el punto de vista económico, tanto para aplicaciones domésticas, como para reproducción en auditorios de mayor dimensión y más público.

1.2. Potencial y alcance

El presente trabajo trata primordialmente sobre un conjunto de programas para el procesamiento de señales de audio que se han implementado sobre un arreglo lineal de altavoces y que logran agregar efectos diversos a las señales procesadas. Los más importantes de estos efectos son la lateralización de la señal reproducida, queriendo decir con esto que la señal parece emitirse con un cierto ángulo respecto a la normal del arreglo.

A partir de estos efectos de lateralización, se ha trabajado en la mejora al sistema de reproducción estereofónico (de dos canales) para la extensión de la zona de audición (sweet spot). Para ello se aplica el efecto de lateralización para cada canal de la mezcla estéreo de modo que a un canal se le confiere una lateralidad hacia la derecha y al otro canal hacia la izquierda. De este modo, se tiene que en los diferentes puntos de la zona de audición la señal llega con un ángulo dado, para la señal de cada canal, haciendo que se presente una situación análoga a la de la zona óptima de audición de un sistema estereofónico de dos canales.

Asimismo, se ha hecho una investigación sobre la técnica de reproducción de sonido de Síntesis de Frentes de Onda (WFS) para su implementación con señales en tiempo real para reforzamiento de escenas acústicas y generación de fuentes virtuales. Las virtudes y limitaciones de

dicha técnica se hacen notar y se plantea su implementación bajo ciertas limitantes prácticas.

Una de las principales virtudes de los programas realizados es que su implementación está hecha de tal manera que puede fácilmente ser trasladada sobre cualquier plataforma y con equipo de cómputo que no necesariamente sea de alto rendimiento. Actualmente se han implementado en lenguaje C y con el uso de las bibliotecas portaudio, audiofile y fftw-3 que se distribuyen actualmente como software libre, lo que hace que dichos programas puedan ser utilizados en aplicaciones donde la economía sea un requerimiento.

1.3. Contenido de la tesis

La presente tesis se divide en 5 capítulos en los que presentan la información sobre la investigación hecha y el trabajo realizado durante su realización. El primer capítulo presenta una visión introductoria sobre los sistemas de reproducción de audio, sus objetivos, sus virtudes y limitaciones. Asimismo se presenta un breve panorama del trabajo realizado, con algunas de sus perspectivas y limitaciones.

El segundo capítulo es un resumen de la investigación realizada. Primeramente se habla del sistema de audición biaural humano, las diferentes claves que nos permiten escuchar de la manera que lo hacemos y su relación con la localización espacial, lo que tiene una importancia fundamental en las aplicaciones de reproducción de sonido. Posteriormente se trata de las diferentes técnicas de reproducción sonora existentes con atención a sus logros y limitaciones; asimismo se hace una comparación entre algunas de ellas. Por último se habla de la técnica de Síntesis de Frentes de Onda (WFS) así como de sus logros, potencial y limitaciones.

El tercer capítulo es una descripción de los programas implementados y las ideas bajo las cuales están implementados. Comienza con un repaso de la teoría del arreglo lineal de fuentes acústicas dado que la presente implementación se basa en un arreglo de altavoces. Se presentan simulaciones hechas en Matlab de configuraciones análogas a las implementadas con lo que se pretende hacer una comparación de los resultados respecto a comportamientos ideales. Se describe el equipo y el software utilizados y la ventaja de la portabilidad de esta implementación a otras plataformas.

El cuarto capítulo presenta los resultados de las mediciones hechas para verificar el funcionamiento de los programas en cuanto a la direccionalidad que se pretende dar al arreglo de altavoces. Se describe el arreglo de medición y el equipo utilizado, así como los parámetros y elementos de análisis de señales que permiten obtener información de las

mediciones realizadas. Se presenta y se describe lo que se observa a partir del análisis.

El quinto capítulo presenta las conclusiones a las que se ha llegado a partir de la investigación realizada, el sistema diseñado y su implementación y las mediciones realizadas para verificar su comportamiento. Se ha agregado un apartado con algunas perspectivas de trabajo futuro con respecto a este proyecto y la viabilidad que podría tener esto. Finalmente en el apéndice se encuentra el código de los distintos programas realizados para la obtención, reproducción y análisis de señales.

2. Antecedentes

2.1. Sistema de audición binaural

El sistema auditivo humano es un sistema de audición binaural por naturaleza, dado que cuenta con dos oídos para analizar los sonidos que percibe el individuo. Para ello, se basa principalmente en la evaluación de diferencias en las características de las señales sonoras recibidas en cada oído. Dichas diferencias le permiten al individuo obtener información de los eventos acústicos de su entorno. De esta manera, podemos llevar a cabo funciones muy especializadas, tales como el reconocimiento de interlocutores, la localización de fuentes sonoras, así como la identificación de ellas.

2.1.1. Localización angular

El sistema auditivo humano tiene una capacidad única para localizar e identificar sonidos en un ambiente tridimensional, con notable exactitud. Los escuchas humanos pueden detectar diferencias de tiempo de arribo de alrededor de 7 μ s. La percepción de sonido está basada en una multiplicidad de indicios o indicadores acústicos, que incluyen diferencias de tiempo y amplitud y efectos de respuesta en frecuencia dependientes de la dirección, causados por reflexión de sonido en el pabellón de la oreja, la cabeza y torso, reunidos en lo que se conoce como Funciones de Transferencia Relacionada a la Cabeza (HRTF, por sus siglas en inglés); una para cada oído (izquierdo y derecho) y para cada posible ubicación espacial de una fuente sonora en el espacio libre tridimensional [11].

El proceso de audición humano está basado en el análisis de las señales acústicas que llegan a los dos oídos para diferencias de intensidad, tiempo de arribo y filtraje direccional por el oído exterior. Varias teorías fueron propuestas tan temprano como 1882, una de ellas por S. P. Thompson, que identificaban dos mecanismos básicos de percepción para la localización de una fuente: las diferencias interaurales de tiempo (ITD's) y las diferencias interaurales de nivel (ILD's). [11]

Posteriormente surgió una teoría, propuesta por Lord Rayleigh [11], basada en la combinación de ambas diferencias que operaban para diferentes intervalos de frecuencias y longitudes de onda. Para longitudes de onda cortas (frecuencias de 4-20 kHz) la cabeza proyecta una sombra acústica dando un nivel más pequeño en el oído más lejano de la fuente (ILD, figura 2.1 b). Para longitudes de onda grandes (frecuencias de 20 Hz – 1 kHz) la cabeza es pequeña comparada con la longitud de onda y la localización está basada en las diferencias percibidas en el tiempo de

arriba del sonido en los dos oídos (ITD, figura 2.1a). Los dos mecanismos de diferencias interaurales de tiempo e intensidad, forman la base de lo que se conoce como teoría dúplex de la localización de sonido. En el intervalo de frecuencias medias, ambos mecanismos están activos, lo que en ocasiones resulta en situaciones con varios indicios conflictivos o ambiguos que pueden causar errores de localización.

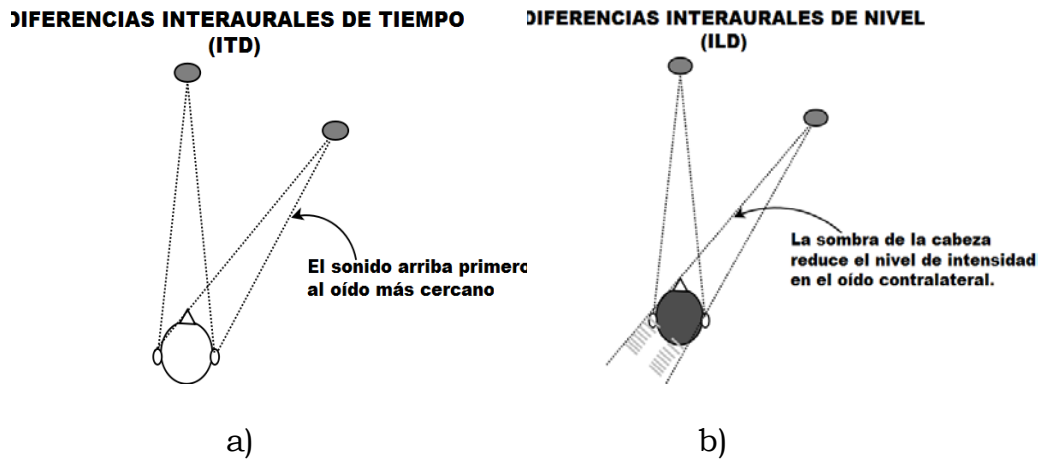


Figura 2.1 Diferencias interaurales de tiempo **a)** (ITD's) y de nivel **b)** (ILD's). El sentido del oído aprovecha estas diferencias para localizar sonidos en el espacio para diferentes intervalos de frecuencia y longitud de onda.

Mientras que las diferencias de intensidad o tiempo proveen información de la dirección de la fuente en el plano horizontal (azimutal), en el plano medio (normal), las diferencias de tiempo y de nivel son nulas y la localización está basada en filtraje espectral. La reflexión y difracción de las ondas sonoras de la cabeza, torso, hombros, y barbilla combinadas con resonancias causadas por el canal auditivo forman las bases de las HRTF. El oído externo puede ser modelado como un sistema lineal e invariante en el tiempo, LTI, que está caracterizado completamente en el dominio de la frecuencia por las HRTF. Como Blauert describe [24], el papel del oído externo es superponer distorsiones lineales en distancia y ángulo sobre la señal de sonido incidente. La información espacial es entonces codificada en las señales recibidas por los tímpanos a través de una combinación de filtros dependientes e independientes de la dirección. La magnitud y fase de esas funciones de transferencia (HRTF) varía significativamente para cada dirección y también de persona a persona.

Además de tales indicios direccionales, hay una multiplicidad de otros indicios, como percepción de timbre, respuesta en frecuencia e intervalo dinámico. Aún más, hay numerosos aspectos subjetivos que varían de escucha a escucha, que son igualmente importantes. Estos incluyen atributos tales como el tamaño aparente de la fuente, la envoltura, la claridad y la calidez.

Para tonos puros, la localización angular es producida parcialmente por diferencias de fase entre los dos oídos, causadas por la diferencia en distancia de la fuente a cada uno de los oídos conforme cambia el ángulo. El efecto de fase, sin embargo, se vuelve ambiguo arriba de 1000 Hz, debido a que las longitudes de onda cortas pueden resultar en que para más de un ángulo haya la misma diferencia de fase. Sin embargo, en la región de más alta frecuencia las diferencias de intensidad producidas por difracción o efectos de sombra de la cabeza, la barbilla y los oídos externos, son ya lo suficientemente grandes como para permitir la localización angular.

Para sonidos complejos, los efectos equivalentes son el tiempo de arribo y la diferencia de calidad o timbre sonoro. Un pulso de onda complejo tiene un frente de onda inicial que arriba en el oído más cercano poco tiempo antes que al oído más lejano. Es esta pequeña diferencia de tiempo la que es aprovechada por el sentido del oído para determinar pequeñas variaciones angulares, particularmente para sonidos que se originan por fuentes cerca del plano medio (directamente adelante). Es típico voltear hacia la fuente para localizarla con máxima precisión.

Para sonidos impulsivos tales como voz o clicks, diferencias tan pequeñas como 1 o 2 grados pueden ser percibidas. Esos ángulos corresponden a diferencias en tiempos de arribo de aproximadamente 10 a 20 μs y la máxima diferencia posible para una fuente en línea con los dos oídos es de sólo 700 μs aproximadamente. La diferencia de amplitud para ángulos tan pequeños es despreciable y debe ser asumido que las diferencias en tiempo de arribo son los que dan las claves para la localización. Por otra parte, no es posible para un solo oído distinguir tales intervalos de tiempo tan cortos. Esta decodificación de diferencias de tiempo de arribo requiere dos oídos, y es lograda gracias al procesamiento neuronal que ocurre en el cerebro.

El efecto del tiempo de arribo es auxiliado por las diferencias de calidad o timbre, causadas por difracción del sonido en los oídos. La diferencia de calidad es otra manera de decir que un cambio en la forma de onda ha ocurrido. Las diferencias de intensidad debidas a difracción son funciones de la frecuencia y causan que un sonido complejo tenga una composición diferente de frecuencia e intensidad en cada oído. Este efecto permite evitar hasta cierto punto, ambigüedades en la localización de dirección, que resultarían sólo del tiempo de arribo, debido a que los efectos de difracción son tan complicados, que para una diferencia de calidad dada, corresponde sólo una dirección. Las diferencias de calidad también cambian rápidamente cerca de la dirección media; por lo tanto, la localización angular es mucho menos precisa en los lados que en frente o detrás. Los cambios tanto en tiempo de arribo como en calidad son relativamente pequeños conforme la fuente se eleva en frente de un observador. Por lo tanto, la habilidad de distinguir ángulos en la dirección vertical es relativamente pobre [8] [11] [12] [22] [23].

2.1.2. Localización de profundidad

Percibir la posición de una fuente sonora en el espacio incluye la determinación de la distancia, así como del ángulo. Debido a que el oído no tiene un mecanismo correspondiente al del ojo para enfocar la fuente, entonces debe depender de claves menos definidas. En ausencia de reverberación, la única información disponible es intensidad y calidad. A partir de experiencia pasada, el sentido del oído permite formarse una idea aproximada de la distancia mediante la interpretación de la intensidad absoluta de un sonido y el juicio de las diferencias de calidad producidas por absorción atmosférica. Esas comparaciones son hechas contra una imagen mental de lo que el sonido debería ser. En presencia de reverberación, el oído puede juzgar distancia basándose en la razón de sonido directo y reverberante. Debido a que ninguno de esos métodos es preciso, el juicio de la distancia es mucho menos preciso que la percepción de ángulo. Probablemente todos hemos tenido la experiencia de haber juzgado mal la distancia de un sonido escuchado por primera vez, mientras que ninguna dificultad fue experimentada en determinar su dirección [12].

2.2. Sistemas de reproducción de sonido

Como ya hemos mencionado, el propósito de un sistema de reproducción de audio es, o debería ser, el de reproducir fielmente un evento acústico, en vivo o grabado o sintetizado -o aún mixto- a una audiencia de un número variable de escuchas, que puede ir desde uno hasta varios cientos o más.

En julio de 1928 en The Gramophone un crítico de música anónimo (K. K.) escribió: "Cuando estoy en una sala acústicamente perfecta (llena de gente), en el mejor lugar para escuchar, y escucho una orquesta, puedo oír tales y tales sonidos. Yo quiero escuchar precisamente este efecto de una grabación, si eso es posible". El director Leopold Stokowski, hablando en la 16ª convención de la Sociedad de Ingenieros de Audio, en octubre de 1964, dijo: "Mediante grabaciones, la radio y televisión estamos atestiguando la más grande presentación de masas de la experiencia musical conocida por el hombre". Aunque más de 80 años han pasado desde que K. K expresó su esperanza y más de 40 desde que Stokowski recordó a la comunidad de ingenieros de audio sus responsabilidades, la meta de la fidelidad perfecta permanece hasta este día como una meta elusiva. Con cada subsecuente mejora en la fidelidad eléctrica de una señal sonora grabada o transmitida, algún progreso (no poco) ha sido hecho, pero K. K se refería específicamente al efecto de cómo él escuchaba [27].

Hoy sabemos que éste es también un asunto psicofísico y que presenta retos técnicos y económicos para la industria del audio sobre los cuales no hay un consenso en el modo de abordarlos, la historia misma del desarrollo de la tecnología de audio está llena de diferentes enfoques y tecnologías, batallas comerciales, etc. Cualquier intento por hacer una evaluación se encara inmediatamente con la cuestión de cuál debe ser la meta de un sistema de reproducción: debe ser transportar al oyente a la escena o si la meta debe ser transportar la escena al recinto del oyente.

El uso de los canales individuales que hoy usamos, viene dado por la tecnología disponible (los micrófonos y altavoces son básicamente dispositivos con ubicaciones específicas en el espacio y los equipos electroacústicos están diseñados para el manejo de un número pequeño de canales de audio individuales). Sin embargo, el sonido es un fenómeno dinámico continuo en tres dimensiones. La cuestión es cuán efectivo es un sistema para la reproducción en términos perceptivos.

La respuesta quizá depende de la naturaleza del material acústico y aún de las preferencias de los oyentes. Hay situaciones donde el objetivo lógico es la aproximación de transportar al oyente a la escena (la música sinfónica, cine, etc.) y otras donde esto no es estrictamente necesario. En la situación actual, debemos estar contentos con que alguna de dichas posibilidades sea realizable.

Existen diferentes sistemas de reproducción, con calidades de pobres a excelentes, dependiendo del material grabado y el medio de reproducción. Una característica de muchos de esos sistemas, es que dependen fuertemente de efectos psicoacústicos, como las llamadas fuentes fantasma o fuentes virtuales, que son fuentes aparentes de sonido que se escuchan entre los altavoces y que dependen de las diferencias de amplitud y tiempo entre las señales de los altavoces.

Desde la invención del fonógrafo, las técnicas de grabación y reproducción sonora han evolucionado constantemente. Uno de los temas importantes en la investigación siempre ha sido preservar las características espaciales y temporales de los eventos sonoros grabados. En este sentido, los sistemas estereofónicos de dos canales fueron introducidos en el mercado en la década de 1960 para, por primera vez, añadir la dimensión espacial a la reproducción de sonido a nivel de consumidor. El siguiente gran avance en este tema fue el paso hacia la reproducción de campos sonoros que intenta recrear las variaciones de presión que permiten a un escucha percibir la posición de fuentes sonoras viniendo de localizaciones arbitrarias en el espacio tridimensional [21].

La reproducción del campo sonoro puede ser generada esencialmente de dos maneras distintas. La primera es mediante proveer señales directamente en los oídos, las cuales son similares a las que deberían ocurrir si fuentes sonoras reales estuvieran en las posiciones deseadas. Esta aproximación se conoce como reproducción binaural

(cuando se efectúa mediante audífonos) o transaural (cuando se efectúa mediante altavoces). La segunda manera es recrear el campo sonoro sobre una región del espacio más extendida, que será interpretada por el escucha como estar en el ambiente que se desea simular. Esta aproximación es una reproducción basada en holografía. Uno de los primeros sistemas holográficos codificaba información y fue propuesto por Gerzon (1973) y se desarrolló más tarde como el sistema Ambisónico. Una aproximación alternativa es el sistema conocido como Generación de Frente de Onda [Wave Field Synthesis (WFS)] [21].

WFS es un técnica de reproducción sonora espacial que genera un campo sonoro real utilizando arreglos de altavoces. Los campos sonoros son generados basándose en fuentes virtuales en la escena detrás de los altavoces o aún en el área de escucha (sección 2.3). Cuando se utiliza reproducción sonora basada en WFS los campos sonoros pueden ser generados de manera correcta espacial y temporalmente. Por lo tanto, los oyentes experimentan la sensación de que el origen del sonido está realmente en la posición de las fuentes virtuales. Más aún, el campo sonoro sintetizado es correcto para un área de escucha extendida, con dimensiones mucho mayores que el punto de reproducción óptima (o “sweet spot”) de los sistemas surround comerciales [18] [20] [21].

2.2.1. Monoaural

Durante la primera mitad del siglo XX, prevaleció el sistema monoaural, que es un sistema que utiliza un canal de grabación y uno de reproducción. Su uso motivó el desarrollo de las técnicas de capturar sonido por medio de micrófonos. Hoy en día, es fácil deplorar este sistema por su falla para representar cualquier aspecto espacial. Aún así, hay situaciones en las que el uso de este sistema debería ser suficiente, por lo menos en teoría (reproducción de solistas o fuentes localizadas en ambientes relativamente libres de ecos). Por otra parte, el sistema monoaural no tiene problemas de doble arribo e imágenes fantasmas y asimismo, es un sistema que está libre del problema de posicionamiento óptimo (“sweet spot”) que tienen los sistemas multicanal [45].

2.2.2. Biaural

Se trata de un sistema que emplea dos micrófonos para grabación, preferiblemente en una cabeza artificial, dos canales independientes y dos audífonos para cada escucha. Esto duplica la audición normal. La reproducción biaural como se define aquí, es la audición ordinaria con dos oídos, debido a que el sistema reproductor transmite una copia fiel del sonido original a los oídos del escucha.

Cuatro años después que Thomas Alva Edison presentara su famosa patente para una “Mejora en fonógrafo o máquinas parlantes” una demostración de una transmisión biaural fue presentada en La Lumière Électrique; la exhibición tuvo lugar en la Gran Ópera de París en 1881

[27]. Diseñado por Clement Ader, el sistema empleaba cuatro pares de micrófonos conectados a receptores telefónicos binaurales en cuatro salas de escucha remotas en el Opera House. Aunque la distancia entre los micrófonos de cada par era aproximadamente un metro, un oyente observó que “al escuchar con dos oídos en los dos teléfonos, el sonido toma un carácter especial de relieve y localización, el cual un solo receptor no puede producir.”

Diez años después, en los Estados Unidos, la American Telephone and Telegraph Company también condujo experimentos con conciertos telefónicos entre Newton, Massachusetts y la ciudad de Nueva York. Años más tarde, la compañía telefónica finalmente demostró transmisiones de sonido multicanal a larga distancia en colaboración con el director Leopold Stokowski. Stokowski, quien se convirtió en miembro honorario de AES y un orador en las convenciones, fue considerado el primer operador de tableros estéreo [27].

En 1933, Fletcher y sus colaboradores de los Laboratorios Bell, presentaron una demostración de tres canales que transmitía las señales de micrófonos de la orquesta de Filadelfia, tocando en la Academia de Música en Filadelfia, a tres altavoces correspondientes en la Sala Constitución en Washington. Stokowski no condujo en esa ocasión, sino que operó los niveles de audio del equipo. Esa no fue la única colaboración de él con los investigadores de los laboratorios Bell. En 1936, para un concierto al exterior en el Tazón Hollywood, él insistió en que Fletcher instalara un sistema de reforzamiento de audio de tres canales, en lo que fue el primer sistema de este tipo, pero ciertamente no el último [1] [11] [12] [27].



a)



b)



c)

Figura 2.2 Colaboración del director Leopold Stokowski al audio con Harvey Fletcher. **a)** y **b)** verificando equipo de reproducción **c)** Stokowski regulando los niveles en el concierto de 1936 en Los Angeles, bajo la supervisión de Fletcher.

Las grabaciones binaurales intentan reproducir con precisión en cada oído del escucha, la presión sonora generada por un conjunto de fuentes, así como sus interacciones con el ambiente acústico. Tales grabaciones pueden ser hechas con micrófonos de medición especialmente diseñados, que están insertados en los canales auditivos o utilizando un maniquí diseñado para esto. Los sonidos grabados utilizando métodos binaurales, son entonces reproducidos a través de audífonos que envían el sonido deseado a cada oído. Las limitaciones asociadas a estas técnicas son:

- a) No existe información de HRTF's individualizada para cada escucha y utilizar algún promedio resulta inadecuado para unificar la percepción de sonido para cada individuo.
- b) Hay grandes errores en la percepción de posición de sonido asociada con el uso de audífonos, especialmente en frente.
- c) El uso de audífonos resulta incómodo para periodos de tiempo largo.
- d) Es difícil evitar la sensación de sonido "dentro de la cabeza".

El uso de altavoces para la reproducción bajo estas técnicas puede sortear dichas limitaciones. Para enviar la señal binaural apropiada para cada campo sonoro, es necesario eliminar el efecto de cruce ("crosstalk") que es inherente a todos los sistemas basados en altavoces. El efecto de cruce viene del hecho de que mientras cada oído recibe el sonido deseado también recibe el sonido "indeseado" del altavoz del lado opuesto. Varios esquemas han sido propuestos para lograr la cancelación de efecto de cruce y se basan en el pre-acondicionamiento de la señal en cada altavoz,

de modo que se genere el sonido binaural deseado en cada oído. Sin embargo esto resulta en un área de audición muy limitada [11] [12] [21].

2.2.3. Estereofónico



a)



b)

Figura 2.3 El ingeniero Alan D. Blumlein **a)** trabajando con un sistema estereofónico en colaboración con R. Beardsley. **b)** Blumlein en conferencia en IEEE 21 de abril de 1938.

Este ha sido el sistema comercial de uso más común desde la década de 1960. La atracción intuitiva del estéreo es evidente: tenemos dos oídos y el sistema de sonido estereofónico provee dos canales correspondientes. Se ha dicho que la historia del audio puede ser dividida en dos periodos: antes y después de Blumlein y su sistema estereofónico. La patente de 1933 del ingeniero británico Alan Dower Blumlein fue un hito [25], aunque la tecnología que él propuso, no pudo ser puesta en práctica sino hasta veinte años después; las demandas que la Segunda Guerra Mundial impuso en los esfuerzos de los científicos e ingenieros crearon una distracción.

Un elemento importante de la reproducción estereofónica es la creación de imágenes virtuales, que son sonidos que “aparecen” en el espacio entre los dos altavoces. Estas imágenes surgen a partir de dos características importantes para el sistema auditivo humano de localización horizontal: diferencias interaurales de tiempo y de amplitud o intensidad sonora.

Estos dos aspectos, además de darle cualidades espaciales, le dan al sistema también una alta dependencia de la posición del escucha (“sweet spot”). También se presenta el problema del doble arribo (“crosstalk”): el oído derecho escucha una contribución del canal izquierdo y viceversa. También es difícil crear imágenes virtuales fuera del ángulo de escucha delimitado por los altavoces.

Pese a ello, el sistema estereofónico fue un gran avance como lo muestran los eventos históricos que siguieron desde su aparición y aún hasta nuestros días. Además de la industria de las grabaciones, la industria de la radiodifusión, el cine y otras, siguieron pronto la línea del

sonido estéreo, desde su comercialización, cuando en 1961 la FCC autorizó la transmisión de radio FM en estéreo, adoptando los estándares técnicos basados en la invención de Carl Eilers de Zenith, mientras que rechazaron otra, quizás igual de viable de Murray Crosby [27].

La transmisión de radio AM en estéreo fue también tomada en cuenta, por lo menos 2 veces. Cinco sistemas estéreo fueron propuestos a la FCC en el mismo expediente que consideró los estándares para el FM estéreo, pero dicha comisión declinó aceptar alguno. Otro intento ocurrió en 1978 con la propuesta a la FCC del reporte final del Comité Nacional de Radio AM Estereofónico (NAMSRC), pero el comité fue incapaz de recomendar un sistema específico, por lo que se decidió dejarlo como una decisión de mercado. El mercado aparentemente decidió no adoptar un sistema de sonido estéreo en el AM.

Aunque muchos de los principios del sonido estereofónico fueron desarrollados a través de resultados de investigaciones en los primeros años de la década de 1930, aún hay un malentendido del significado de la palabra estéreo misma. La palabra estéreo, del griego *stereos*, significa sólido (y por extensión semántica: tridimensional). Mientras que generalmente se asocia este término con la reproducción en dos altavoces, la reproducción estéreo no está limitada a ello. Los famosos experimentos de los Laboratorios Bell sobre estereofonía en los años de 1930 utilizaron tres canales. El estéreo cinematográfico en los años de 1950 usaba no menos de cuatro canales y tantos como siete. Aunque cuando el estéreo finalmente se volvió de uso doméstico en 1958, sólo fueron usados 2 canales. Esto se debió no a las preferencias de los oyentes, ni por predisposición de parte de los profesionales del audio, sino que dos canales eran los que el fonógrafo de aquel entonces podía grabar y reproducir [11] [39].

La estereofonía comenzó con el trabajo de Blumlein en Gran Bretaña, quien pronto reconoció que era posible localizar un sonido dentro de un intervalo angular utilizando una combinación apropiada de retrasos y diferencias de nivel. Fletcher, Steinberg y Snow en los Laboratorios Bell, tomaron una aproximación diferente. Ellos consideraron una “pared de sonido”, en la cual un número infinito de micrófonos es usado para reproducir un campo sonoro a través de un número infinito de altavoces, en analogía al principio de Huygens. Ellos mostraron que un sistema de tres canales, consistiendo de canales izquierdo, derecho y central en el plano azimutal, podía representar la lateralización y profundidad del campo sonoro deseado con aceptable exactitud. El primer sistema estereofónico de tres canales fue demostrado en 1934 con la Orquesta de Filadelfia interpretando remotamente para una audiencia en Washington D. C., sobre una línea telefónica de banda ancha [1] [21].

2.2.4. Cuadrafónico

Mientras que los métodos estereofónicos pueden ser una herramienta poderosa para la reproducción de atributos espaciales de un campo sonoro, fallan pronto en la reproducción tridimensional real. El sistema de reproducción cuadrafónico intentó sortear tales limitaciones mediante el propósito de capturar y transmitir información de los campos sonoros directo y reverberante. Para entregar cuatro canales sobre un medio de dos canales (el fonógrafo de aquel entonces) fue necesario desarrollar un esquema apropiado de codificación y decodificación. Varios esquemas fueron propuestos basados en la manipulación de las señales estéreo originales. Dichos sistemas fueron capaces de reproducir imágenes sonoras con bastante precisión en el frente y detrás del plano azimutal; sin embargo, exhibían serios problemas cuando intentaba reproducir imágenes sonoras al lado del escucha. Los experimentos mostraron que ésta era una limitación asociada tanto a la síntesis de sonido con cuatro canales, como a los mecanismos psicoacústicos humanos. Esas limitaciones técnicas y el hecho de que existían dos formatos competidores en el mercado, contribuyeron a un temprano desuso de los sistemas cuadrafónicos.

En 1968 un joven profesional ejecutante de fagot puso a la comunidad de ingeniería de audio en un curso que consumió casi todas sus energías durante la siguiente década. Esto inició batallas apasionadas en la industria del mercado y los comités técnicos aún más encarnizados que la famosa “guerra de las velocidades” de la industria de los discos sobre la velocidad de 45 o 33 rev/min. Peter Scheiber, el fagotista no fue el primero en sugerir que el audio de cuatro canales parecía ser buena idea. Pocos años antes, los productores Joanna Nickrenz y Marc Aubort comenzaron a comercializar grabaciones en cinta de cuatro pistas y David Haffler, un violinista que se volvió ingeniero, diseñó un circuito matricial que conducía los componentes fuera de fase de una señal estéreo a un par de altavoces. Sin embargo, Scheiber fue más allá, él mostró que 4 canales análogos podían comprimirse en dos y restaurarse útilmente a 4 (una codificación en matriz 4-2-4) si eran empleadas redes de desfase de banda ancha en los procesos de codificación y decodificación [11] [27].

Las primeras demostraciones de Scheiber generaron en la industria generalmente una recepción tibia e indiferente, pero una dada a los ejecutivos de CBS llamó la atención de Benjamin B. Bauer. Respaldado por los recursos de los Laboratorios CBS, Bauer creó su propia matriz, llamada SQ (por estéreo-Quad) y para la cual él atribuía características espaciales superiores y compatibilidad con la reproducción estéreo o monofónica.

Bauer no tuvo el campo para sí mismo durante mucho tiempo, pronto hubo una matriz SQ de la Corporación Sansui, QMX de la Universidad de Illinois y Ambisonics de Michael Gerzon en el Reino Unido

y un día ya no fue una cuestión simple cuál matriz debía ser superior, sino, mas bien, si las grabaciones del futuro deberían estar basadas en una matriz o en un formato discreto. Además un joven ingeniero de California, Luis Dorren, se había convencido de la cuadrafonía y había inventado una mejora a los estándares de transmisión de FM estéreo, el cual afirmaba tener recepción discreta de 4 canales [47].

Como con el trabajo de Bauer, el trabajo de Dorren atrajo la atención de otros, y cuando la Asociación de Industrias Electrónicas (EIA) estableció su Comité Nacional de Radio Cuadrafónico (NQRC), en 1972, tenía trabajos de Dorren, RCA, Cooper Nippon Columbia, General Electric, Zenith y CBS, los últimos demandando sólo el uso de la matriz SQ para señal estéreo de banda base, mientras que todos los otros empleaban varias sub-configuraciones de sub-portadoras. El NQRC envió su reporte final a la FCC en 1975, esperando obtener la pronta modificación de las reglas de radiodifusión, pero el tema palideció en la FCC por una década sin acción por parte de los reguladores. Cuando la decisión finalmente fue tomada, esta consistió simplemente en permitir el uso de sub-portadoras adicionales en la transmisión de FM para cualquier propósito deseado. El sistema cuadrafónico probó tener virtudes en la ubicación de fuentes frontales y traseras, pero se vio envuelto sobre todo en una serie de batallas comerciales, además de sus propias limitaciones, que lo hicieron inviable.

2.2.5. Sistemas envolventes (surround)

En los últimos 10 años, un sistema de 5 altavoces ha aumentado notablemente su popularidad (surround 5.1). El oyente es rodeado por altavoces y se asume que son producidas percepciones más realistas. El formato 5.1 está pensado para hacer frente a los defectos del sistema estereofónico. Este formato emplea tres altavoces frontales, dos altavoces laterales (surround) y uno más (0.1) para la reproducción de bajas frecuencias. Este sistema mitiga significativamente los problemas estereofónicos de las imágenes virtuales y los altavoces surround proveen una experiencia más inmersiva que la estereofonía de sólo dos canales. Hay también otros estándares para posicionamiento de altavoces, los cuales utilizan más altavoces alrededor del oyente. Algunas de esas configuraciones tienen un elevado número de altavoces, así pues se habla también de sistemas 7.1, 9.1, 10.1, 22.2, etc. [36] [11] [17] [40].

Un gran hito en la tecnología del sonido multicanal llegó en 1940 cuando los Estudios Disney lanzaron Fantasía, una película con extractos musicales dirigidos por Stokowski y que eran acompañados por imágenes concebidas en Disney. El filme empleaba un nuevo proceso llamado "Fantasound" y que fue desarrollado por W. E. Garrity y J. N. A. Hawkins en los Estudios Disney Hyperion [27] [48]. Los primeros experimentos con Fantasound empleaban 3 altavoces en el frente de la escena y dos altavoces adicionales en las esquinas traseras del auditorio. Dos pistas de audio fueron grabadas ópticamente, una alimentando el altavoz central y

la otra para los otros cuatro altavoces. Un técnico de mezcla proveería el control manual de los niveles de reproducción y la perspectiva acústica durante la ejecución del filme. Esto, sin embargo, prontamente resultó impráctico, debido a que era demasiado complicado para que lo manejara una sola persona. Fantasía tuvo su premier en el Teatro Broadway en Nueva York, en 1940.

Estos sistemas pueden evocar imágenes sonoras localizadas con mucha precisión, que están primariamente confinadas en el plano horizontal, y sonido difuso a los lados y detrás del escucha. Además del canal izquierdo y derecho, estos pueden hacer uso de un altavoz central que ayuda a crear una imagen sólidamente anclada al centro, así como dos altavoces para el sonido ambiente.

En los primeros años de la década de 1950, el primer formato de sonido multicanal fue desarrollado por la compañía cinematográfica 20th Century Fox. Un filme estereofónico fue reproducido sobre tres altavoces frontales. Esos nuevos formatos incluían un canal monofónico adicional que era reproducido detrás de la audiencia y que se conoció como canal de efectos. Este canal incrementó el sentido de espacio para la audiencia, pero tenía una seria limitación tecnológica: los escuchas en medio de los altavoces percibían el sonido dentro de la cabeza y los que estaban cerca de uno, percibían el sonido como proveniente de ese altavoz. La solución fue introducir un segundo canal y reproducirlo sobre un arreglo de altavoces a los lados del teatro para crear un campo sonoro más difuso.

A mediados de 1970, una nueva tecnología sonora fue introducida por los Laboratorios Dolby, llamada Dolby Stereo. Estaba basada en tecnología óptica y sorteaba los problemas asociados con grabaciones de varias pistas magnéticas. Dolby desarrolló un método de matriz para codificar cuatro canales en dos canales [49] [11] [7].

En 1992 Dolby introdujo un nuevo formato llamado Dolby Stereo Digital. Dicho formato eliminaba el proceso de codificación-decodificación y proveía cinco canales discretos en una configuración conocida como estéreo envolvente (Stereo Surround). Un sexto canal para bajas frecuencias fue añadido y prevenía que los altavoces principales operaran con demasiada exigencia a bajas frecuencias. El ancho de banda de dicho canal estaba limitado entre 0 y 120 Hz, que está fuera del rango humano de localización en un cuarto reverberante y que simplificaba los requerimientos de ubicación para el subwoofer utilizado para este canal [11] [49] [50] [51].

Avances recientes en la compresión de audio digital y almacenamiento óptico han hecho posible entregar más de seis canales de audio en un formato de consumidor centrado alrededor del esquema Dolby AC-3. Mientras que hay varios millones de sistemas de consumidor capaces de reproducir más de dos canales, la mayoría de los usuarios encuentra que el uso de muchos altavoces es impráctico.

Los más comunes de estos sistemas incluyen el Dolby Pro Logic y el Dolby Digital, el DTS y MPEG-2. El sistema Dolby Digital 5.1, por ejemplo, fue adoptado como el estándar en E. U. para el sistema de televisión de alta definición. Los requerimientos de diseño de tales sistemas incluyen una cobertura uniforme para la audiencia, localización precisa relativa a imágenes visuales en pantalla, envío de sonidos ambiente difusos y capacidad para la reproducción del amplio rango dinámico presente en las películas [6] [7] [11] [21].

2.3. Sistema estereofónico convencional

El sistema estéreo es un sistema que emplea dos o más micrófonos espaciados en frente de un área de grabación, conectados por canales amplificados independientemente, a dos o más altavoces frente a un área de escucha. Esto crea la ilusión de sonidos que poseen dirección y profundidad en el área entre los altavoces. Aquí aparece una diferencia fundamental respecto al sistema binaural que es preciso hacer notar. El sistema binaural duplica en los oídos del escucha los sonidos que este debería escuchar en el punto de grabación. El sistema estéreo, por otra parte, produce un patrón de sonido anormal en los oídos del escucha que el sentido del oído interpreta como indicativo de dirección en el espacio entre los altavoces. Se dice apropiadamente que el sistema binaural transporta al escucha a la escena original, mientras que la estereofonía transporta la fuente sonora al recinto del escucha.

La reproducción estereofónica ha traído un incremento verdaderamente notable en el realismo del sonido y en el placer de escucharlo, con respecto a las grabaciones monofónicas. En un intento para medir esto cuantitativamente, Fletcher reportó que observadores escuchando alternativamente reproducción monofónica y estéreo (en la reproducción estéreo fueron insertados filtros pasobajas, mientras que la reproducción monofónica fue mantenido un espectro plano hasta 15 kHz) la mitad de observadores aún prefería la reproducción estéreo aun cuando la frecuencia de corte del filtro pasobajas era reducido hasta 5 kHz [12].

Con el objeto de ver por qué el éxito del sistema estéreo, es preciso tomar en cuenta los factores que posibilitan el disfrute de una presentación musical en vivo, cuya emulación debería ser la meta de cualquier sistema de reproducción sonora. Cuando se escucha una ejecución musical en vivo, podemos decir donde están localizados los varios instrumentos y esto es parte del disfrute. La otra característica importante de la audición en la vida real, es la sensación de estar en un recinto; efecto que es producido principalmente por reverberación. Se desea que en un sistema de reproducción sonora, el escucha oiga el sonido viniendo primeramente de la fuente e igualmente reflejado, análogo a las características ya mencionadas de la audición *in situ*.

Se acostumbra describir la reproducción estereofónica ideal como sigue: una pantalla que consiste de un número extremadamente grande de micrófonos extremadamente pequeños es puesta en frente de una fuente sonora. Cada micrófono es conectado a un altavoz correspondiente (extremadamente pequeño) en una pantalla de altavoces puesta ante la audiencia. Entonces el sonido proyectado a la audiencia será una copia fiel del sonido original y un escucha oirá el sonido con verdadera perspectiva.

La miriada de altavoces en la pantalla, actuando como fuentes puntuales de sonido con los sonidos escuchados por los micrófonos proyectarían una copia fiel del sonido original dentro del área de escucha. El escucha emplearía entonces su audición biaural ordinaria y sus oídos serían estimulados por sonidos idénticos a esos que él hubiera escuchado viniendo de la fuente original de sonido. Como se muestra en la figura 2.4 esto significa un pulso directo de sonido a cada oído por cada pulso simple de la fuente.

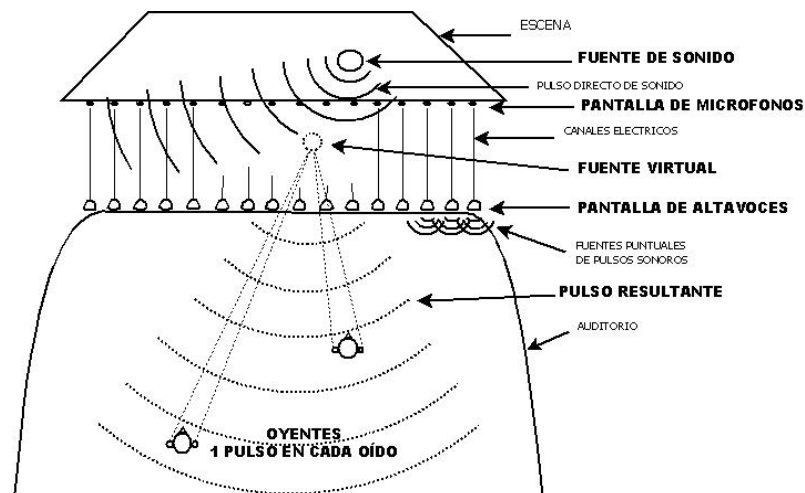


Figura 2.4 Esquema de la implementación del sistema estereofónico ideal. El sonido es grabado mediante una pantalla de muchos micrófonos y reproducido por una pantalla correspondiente de altavoces. El efecto resultante es un pulso que corresponde a una fuente virtual que es una copia de la fuente real grabada.

2.3.1. Condiciones de operación. Ilusión creada.

Está establecido que el uso de la miriada de canales mencionada en la descripción anterior es impráctico e innecesario. Una buena perspectiva acústica puede ser lograda con tan sólo dos o tres canales. La figura 2.5 ilustra las condiciones para un arreglo típico, donde sólo tres canales son usados. Este arreglo da ciertamente buena perspectiva auditiva; pero lo que no es generalmente apreciado, es que las condiciones son ahora muy diferentes de la impráctica pantalla infinita; de modo que un diferente mecanismo de audición es usado por el cerebro.

Cada altavoz envía un pulso al escucha, quien recibe, por tanto, tres copias fieles del sonido en cada oído en rápida sucesión. La diferencia de tiempo entre esos pulsos es tan corta como para no permitir al oído distinguirlos separadamente; consecuentemente, el mecanismo de audición los funde en una ilusión de un simple pulso viniendo de una fuente virtual localizada en alguna parte en el espacio entre los altavoces. Este tipo de audición cae fuera de la experiencia normal, afortunadamente el cerebro es capaz de formar una impresión sencilla y concisa de lo que podría esperarse fuera un conjunto confuso de señales enviadas por los oídos.

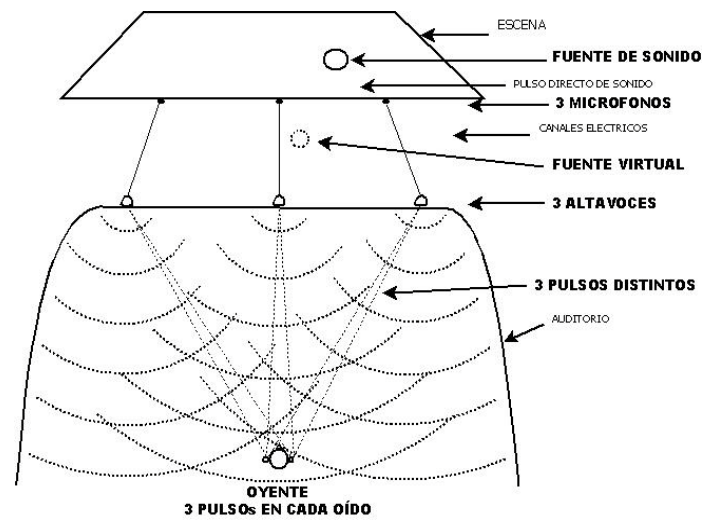


Figura 2.5 Esquema de la implementación del sistema estereofónico de 3 canales. Este arreglo ofrece una reproducción múltiple (3 veces) del sonido original el cual es interpretado por el escucha como un sonido viniendo de una sola fuente virtual (sonido fantasma).

2.3.2. Percepción angular

Diferencias de intensidad. Las diferencias entre los pulsos de sonido directos ocasionan que el escucha tenga la sensación de localización angular de la fuente virtual. La diferencia más obvia entre ellos es la intensidad del sonido proyectado por los varios altavoces. Está bien establecido que las diferencias de intensidad en los canales contribuyen de manera extremadamente importante a la percepción angular. Esas diferencias son causadas por las diferentes distancias de la fuente a los micrófonos durante la grabación. Cuando la fuente se mueve cerca de un micrófono, la salida del altavoz correspondiente es mayor que las de los otros altavoces y la localización va en la misma dirección que la fuente real. El efecto del tiempo de arribo es hacer que la localización tienda hacia el altavoz del cual el pulso llega primero. A esto se le conoce como efecto de precedencia [12] [36] [52] [53] [54]. Este efecto es fuerte para pequeñas diferencias, digamos 2 o 3 ms y tiende a debilitarse para diferencias de tiempo mayores.

Diferencias de calidad. Si los micrófonos tienen variaciones de directividad con la frecuencia hay diferencias de calidad así como diferencias de intensidad en los canales conforme la fuente se mueve. La localización angular estará definitivamente afectada por esto.

Por otra parte, las frecuencias muy bajas no contribuyen esencialmente a la localización. Por ejemplo, si se insertan filtros pasobajas de 1 kHz, los resultados de localización resultan pobres, y por otra parte, si se eliminan frecuencias debajo de 300 Hz, no se produce ninguna diferencia en la localización. Se ha encontrado que mucho del efecto estereofónico se preserva si las frecuencias bajas son reproducidas a partir de una sola unidad de baja frecuencia, y utilizando canales laterales reproduciendo solamente frecuencias arriba de 300 Hz. Esto es de gran valor práctico para la reproducción estereofónica económica, tal como la que requieren los sistemas domésticos de reproducción musical.

Percepción de profundidad.

La percepción de profundidad en la reproducción estereofónica es controlada esencialmente por los mismos factores que en la audición ordinaria: intensidad, calidad y relación de sonido directo a reverberante. Conforme la intensidad de sonido decrece, la impresión producida es la de una fuente de sonido alejándose. La misma ilusión acompaña a la pérdida relativa de altas frecuencias. Sin embargo, la contribución más importante en la percepción de profundidad, es el cambio en la razón de sonido directo a reverberante. Conforme la energía reverberante se vuelve más prominente, la fuente parece alejarse en la escena.

2.3.3. Efecto de la posición del observador

Hasta este punto, hemos supuesto por simplicidad que todas las posiciones del escucha son igualmente buenas. Realmente, esto está muy lejos de ser el caso. Los diversos factores que producen el efecto estereofónico resultan desventajosos en algunos aspectos con respecto a las posiciones del escucha.

Movimiento aparente de la fuente, conforme el escucha se mueve: los efectos descritos anteriormente, caracterizan la audición en la posición de escucha óptima de la figura 2.6 (“sweet spot”) o en otras posiciones de escucha en la línea central entre los altavoces, donde las distancias a los altavoces laterales son iguales. Sin embargo, conforme el observador se aleja de dicha posición central, pueden ocurrir grandes desplazamientos de la fuente virtual. La percepción estereofónica de espacio es preservada, y la fuente virtual continúa moviéndose, pero no está localizada en el mismo lugar de la escena para todos los escuchas, como ocurriría si ellos estuvieran en una escena real.

La figura 2.6 ilustra lo que sucede para una fuente en el centro del área de grabación y una implementación típica. El escucha en la posición central (conocida como “sweet spot”, posición óptima) recibe pulsos idénticos de sonido de los canales laterales. De modo que la imagen virtual se ubica en el centro de la línea que une a los altavoces, esto es lo que se conoce como imagen fantasma. Sin embargo, conforme se mueve a la izquierda, recibe señales diferentes en los oídos. Se ve que el altavoz izquierdo ahora contribuye con más intensidad y con una señal anticipada, respecto a la posición óptima; dichos efectos sabemos ya que hacen tender la localización en esa dirección. Este es el caso, y conforme el escucha se mueve a la izquierda, la posición de la fuente virtual se mueve en la misma dirección.

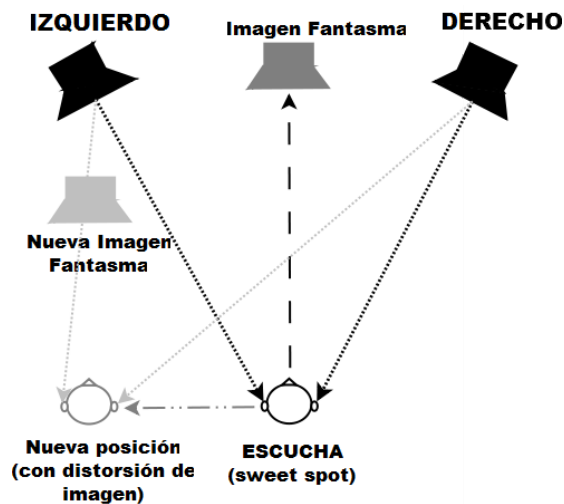


Figura 2.6 Efecto de cambiar la posición de escucha sobre la imagen fantasma. Conforme el escucha se aleja de la línea central de los altavoces (“sweet spot”), el incremento en intensidad y decremento en tiempo de arribo del altavoz más cercano, modifica la ubicación de la fuente virtual hacia el mismo.

Cualitativamente se encuentra que tiene lugar un gran desplazamiento de la fuente virtual para pequeñas desviaciones de la posición del escucha respecto de la posición óptima, donde los cambios relativos de intensidad son pequeños. De modo que dicho desplazamiento debe ser adscrito a cambios en el tiempo de arribo.

Una razón que ayuda a explicar este efecto, es que en la reproducción con dos altavoces, las señales en los oídos del escucha están formadas por la interferencia de ondas acústicas emanando de los altavoces. Si la cabeza se mueve alejándose del punto óptimo, el efecto estéreo se distorsiona, debido a señales inesperadas creadas por interferencia no prevista. Las causas primarias del cambio en la interferencia son diferencias en tiempos de arribo y amplitudes, y debidas a las diferentes distancias a los altavoces y reflexiones secundarias descompensadas. Este problema se vuelve más serio, aún cuando la

distancia entre los altavoces no es grande en comparación con las desviaciones de la posición central, como ocurren en monitores de PC y aplicaciones de TV.

La idea de lograr un incremento en área de escucha (“sweet spot widening”) en una configuración estéreo ha sido introducida y estudiada a principios de la década de 1990, y fue llamada ‘Estéreo Independiente de la Posición’. La principal idea es la siguiente: si el escucha se mueve a la izquierda, la intensidad de sonido del altavoz derecho debe incrementarse mientras que la del altavoz izquierdo debe disminuir, de modo que la imagen sonora virtual permanezca en el medio. Esto puede ser visto como una especie de control automático de balance, dependiente de la posición del escucha. Varios esquemas han sido propuestos [9] [10].

2.3.4. Número de canales

El número de canales dependerá del tamaño de la escena, del recinto donde se escucha y de la precisión de localización deseada. Dos canales dan una gran medida del efecto de “espacio” deseado en la reproducción estereofónica y dará localización precisa para una escena pequeña. Para usos como la reproducción musical doméstica, donde se prefiere una solución económica y la localización precisa de las fuentes no es de gran importancia, con tal de que la percepción de separación de fuentes se preserve, la reproducción en dos canales es de real importancia. El uso de tres canales representa una buena elección económica para escenas y auditorios ordinarios. Puede lograrse una buena precisión en la localización para posiciones de escucha favorables, con resultados razonables para otras posiciones. De hecho, los sistemas surround se basan en tres canales estereofónicos frontales [1] [2] [6] [8] [10] [11] [12].

2.4. Límites y comparación de sistemas de reproducción de sonido

Hay dos clases de limitaciones que dificultan la implementación de sistemas inmersivos de audio. La primera clase tiene que ver con limitaciones fundamentales que vienen de leyes físicas y cuyo entendimiento es esencial para determinar la viabilidad de una tecnología particular respecto de sus límites físicos absolutos. Muchas de tales limitaciones fundamentales no son directamente dependientes de la elección de los sistemas, sino que pertenecen al proceso real de propagación y atenuación en recintos de forma irregular. La influencia del ambiente acústico local se manifiesta en la percepción de atributos espaciales, tales como la dirección y distancia, así como en la percepción de volumen del recinto y tamaño de la fuente. La segunda clase de

limitaciones consiste en restricciones que vienen de consideraciones puramente tecnológicas, según el modo en que sean implementadas.

Las evaluaciones de calidad de reproducción de audio han estado relacionadas tradicionalmente con atributos no espaciales del sonido, tales como el timbre o la distorsión. Sin embargo, el creciente uso de audio multicanal ha motivado el estudio de la percepción espacial del sonido, para comprender cómo los atributos espaciales contribuyen a la calidad del sonido, validez ecológica (entendiendo bajo este concepto que los participantes, en su mayoría reaccionen como si estuvieran bajo condiciones naturales o reales [41]), y preferencia.

Se han llevado a cabo varios estudios para comparar los diferentes sistemas de reproducción sonora. A partir de estos estudios, se desprende que el sistema de reproducción ambisónico se percibe como significativamente más envolvente que el sistema estéreo y el transaural. En lo que respecta al realismo, el sistema estéreo es calificado significativamente más alto que el transaural. La calidad general de los sistemas estéreo y ambisónico es evaluada significativamente más alta que la del sistema transaural.

En lo que respecta a la exactitud de localización ofrecida, la evaluación del sistema ambisónico es significativamente menor que para las técnicas estereofónica y transaural; esto es, es menos “legible” entendiéndose con éste término que la localización de fuentes sonoras implica no solamente la percepción de una fuente virtual sino también la diferenciación entre varias de ellas cuando la señal reproducida pretende poner más de una fuente virtual [2]. Esto es especialmente cierto para sonidos grabados en direcciones laterales. Específicamente, el sistema ambisónico provee un buen sentido de la inmersión y de carácter envolvente, pero una pobre localización, legibilidad de la escena, mientras que la estereofonía y la reproducción transaural proveen una localización precisa y buena legibilidad, pero pobre inmersión y carácter envolvente. Los resultados se resumen en la siguiente tabla.

Técnica De reproducción	Fortalezas	Debilidades
Transaural	Precisa y fácil localización. Buena legibilidad	Pobre realismo y falta de inmersión y de carácter envolvente.
Ambisónica	Fuerte inmersión y carácter envolvente.	Pobre localización y legibilidad.
Estéreo	Localización muy precisa	Falta de inmersión y carácter envolvente.

Tabla 2.1 Evaluación de las técnicas de reproducción estereofónica, transaural y ambisónica para desempeño en localización de fuentes sonoras y carácter envolvente.

2.5. Síntesis de Frentes de Onda: WFS, Wave Field Synthesis.

Generalidades

El fin último de un sistema de reproducción de sonido 3D consiste en recrear las sensaciones acústicas que percibiría un oyente concreto dentro de un entorno acústico particular, pero en otra sala de escucha o con otras condiciones particulares. Los sistemas de sonido estéreo son considerados como la aproximación más sencilla al sonido espacial y se han venido utilizando desde los años de 1960, sobre todo en la música. Desde mediados de los años de 1970 se han venido utilizando, en salas de cine inicialmente y en el hogar en años posteriores, los sistemas de sonido envolvente (surround), que aspiran a proporcionar una mejor sensación espacial que el sistema estéreo, utilizando más canales de reproducción. Sin embargo, estos sistemas sólo logran incrementar la sensación de espectáculo en las proyecciones cinematográficas y no proporcionan una verdadera sensación de sonido 3D. Además, la zona útil de escucha (sweet spot) queda prácticamente restringida al punto central del círculo de altavoces, degradándose la percepción fuera del centro.

Alternativamente, una estrategia mucho más realista es reproducir directamente en los oídos del oyente la señal que escucharía si estuviese en el espacio acústico a simular. La señal de sonido 3D puede ser sintetizada si se conocen las funciones de transferencia acústica referidas a la cabeza HRTF (Head-Related Transfer Functions) del oyente. Sin embargo aparecen diversos problemas para este enfoque (sección 2.1). Existen otros sistemas más avanzados como el Ambisónico o Virtual Surround Panning, que son adecuados para zonas de escucha más o menos restringidas, aunque siempre mayores que en los sistemas transaurales con cancelador de efecto de cruce.

Aunque los sistemas de 5 canales están consolidados hoy en día como el estándar en la reproducción de audio multicanal, los sistemas emergentes basados en la generación del campo sonoro están incrementando la atención que reciben. El más popular de esos sistemas es el de Síntesis de Frentes de Onda, WFS (Wave Field Synthesis), que es un sistema de reproducción espacial capaz de sintetizar un campo acústico en un área extendida mediante arreglos de altavoces, haciendo que sea posible crear una escena sonora con características espaciales correctas para cada oyente. La figura 2.7 muestra una implementación típica de esta técnica.



Figura 2.7 Arreglo lineal típico de altavoces en una sala para reproducción mediante WFS. Esta técnica para su implementación requiere de arreglos de un gran número de altavoces.

El sistema de Síntesis de Frentes de Onda descansa sobre el principio de Huygens para crear frentes de onda artificiales o reforzar los ya existentes. En su favor puede decirse que la calidad del sonido producido con un sistema de este tipo es grande. Una de sus principales desventajas es que requiere de un gran número de altavoces para su funcionamiento y un costo en procesamiento elevado [34].

Mirando hacia atrás en el tiempo, el desarrollo de la reproducción de sonido espacial comenzó en los albores de la tecnología de los altavoces. En 1953 Snow publicó una visión general de las técnicas estéreo y discutió la cortina acústica como la técnica ideal de reproducción estéreo, que aspira a transportar la acústica del lugar de grabación a la sala de reproducción mediante el uso de arreglos de altavoces y micrófonos. Debido a los problemas técnicos de ese tiempo, no fue posible poner estas ideas en práctica. Como compromiso, se aplicó la estereofonía de tres canales, aceptando que el objetivo original de recrear el campo sonoro real no sería logrado completamente. Snow describió este precursor del WFS como sigue: “la miríada de altavoces de la pantalla, actuando como fuentes puntuales de sonido idéntico al sonido escuchado por los micrófonos, deben proyectar una copia fiel del sonido original dentro del área de escucha. El observador debería entonces emplear la audición binaural ordinaria y sus oídos deberían ser estimulados por sonidos idénticos a los que debía haber escuchado provenir de la fuente sonora original” [12].

Más tarde, el intuitivo concepto de cortina acústica fue remplazado por una teoría de ondas bien fundada. A fines de la década de 1980, el concepto de WFS fue desarrollado, principalmente en la Universidad Técnica de Delft, Holanda. El origen de esta teoría fue publicado por Berkhout, en 1987 y 1988, en los artículos: “Applied Seismic Wave theory” [15], y “A Holographic Approach to Acoustic Control” [14]. El término “holografía acústica” fue empleado originalmente, no WFS, por su analogía con la técnica holográfica óptica, y el sistema fue diseñado para ser la herramienta definitiva para sistemas de control acústico en teatros. Esas publicaciones establecieron las bases físicas de la WFS al aplicar

algoritmos conocidos de la sismología al campo de la acústica. El trabajo básico sobre WFS fue continuado por Berkhout.

El concepto de la WFS es comparable a la holografía óptica: primero el campo de onda óptico es grabado sobre un plano y después es recreado por una distribución de fuentes de luz, colocadas sobre ese plano. En holografía sonora, el campo acústico es grabado sobre un plano S dado por un arreglo plano de micrófonos. La reproducción es entonces hecha por fuentes sonoras secundarias, altavoces excitados individualmente. [13] [14] [15] [21]

2.5.1. Integral de Kirchhoff-Helmholtz

De la teoría general de la acústica lineal se sabe que un campo acústico arbitrario, encerrado por un volumen arbitrario, puede ser generado mediante una distribución de monopolos y dipolos en la superficie del volumen, con la restricción de que no existan fuentes acústicas dentro de dicho volumen. Esto puede ser representado matemáticamente mediante la llamada integral de Kirchhoff-Helmholtz; ver más adelante, ecuación (2.2).

La teoría de WFS está relacionada al principio de Huygens formulado en 1678. Este principio establece que cada elemento de un frente de onda propagándose en un medio particular puede ser visto como el centro u origen de una nueva onda esférica individual. Consecuentemente, el frente de onda generado por una fuente primaria de sonido puede ser visto como una serie de fuentes secundarias elementales. Ahora bien, no es muy práctico colocar las fuentes acústicas sobre los frentes de onda para la síntesis. También es posible lograr el mismo efecto al colocar los altavoces sobre una superficie (o curva, en el caso de reproducción aproximada bidimensional comúnmente utilizada en WFS) arbitraria fija y ponderar y retrasar las señales de excitación en cada una de ellos un frente de onda acústico puede ser sintetizado mediante un arreglo de altavoces como ilustra la figura 2.8.



Figura 2.8 Ilustración del objetivo fundamental de la técnica WFS. a) fuente sonora original y su campo acústico; b) campo sintetizado mediante el arreglo de altavoces y la fuente virtual correspondiente.

Matemáticamente la formulación de fuente de la Integral de Helmholtz expresa la posibilidad de que la presión dentro o fuera de una superficie pueda ser determinada [13] [29] [30] [33] [38]:

$$P(r) = \iint_S \left[G \frac{\partial P(r)}{\partial n} - P(r) \frac{\partial G}{\partial n} \right] dS \quad (2.1)$$

donde S es una superficie arbitraria cerrada, V es el volumen contenido en S, η es la normal a dicha superficie, r es el vector coordenado de un punto de observación, r_s es el vector coordenado de las funciones del integrando sobre la superficie S (ver figura 2.9) y G es la función de Green de espacio libre dada por:

$$G(r | r_s) = \frac{1}{4\pi R} e^{jkR}$$

en esta última ecuación se tiene que: $R = |r - r_s|$

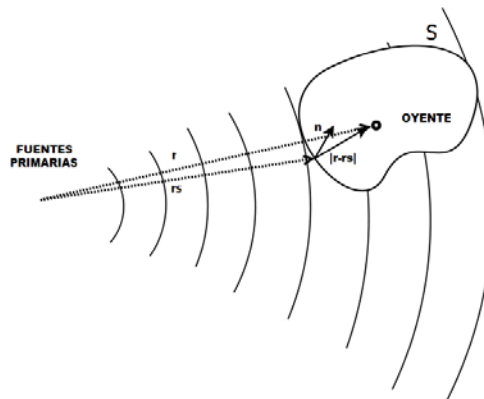


Figura 2.9 Geometría utilizada para la formulación matemática de las integrales de Helmholtz y Kirchhoff-Helmholtz utilizadas en la implementación de WFS.

La ecuación (2.1) establece, considerando el problema anterior, que el campo acústico en el volumen V, generado por los eventos fuera de la superficie S puede ser generado al remplazar esos eventos por una distribución superficial de monopolos y dipolos simples (ver 7.2 en el apéndice), y sumando sus contribuciones sobre la superficie S. De este modo, un campo acústico arbitrario puede ser recreado dentro del volumen V, sin fuentes, mediante fuentes sonoras secundarias distribuidas sobre la superficie cerrada S. Lo último es expresado por la así llamada Integral de Kirchhoff-Helmholtz [13] [15] [21]:

$$P(r, \omega) = \oint_S \frac{1}{4\pi} \left[P(r_s, \omega) \frac{\partial}{\partial n} \left(\frac{e^{-jk|r-r_s|}}{|r-r_s|} \right) - \frac{\partial P(r_s, \omega)}{\partial n} \left(\frac{e^{-jk|r-r_s|}}{|r-r_s|} \right) \right] dS \quad (2.2)$$

Donde $P(r,\omega)$ es la transformada de Fourier de la presión sonora $p(r,t)$, k es el número de onda, ω es la frecuencia angular, r es el vector coordenado de un punto de observación y r_s es el vector coordenado de las funciones del integrando sobre la superficie S (ver figura 2.9) La primera parte de esta expresión representa una distribución de dipolos con la amplitud de fuente dada por la presión sonora, medida sobre la superficie S . El segundo término representa una distribución de monopolos, cuya amplitud es dada por la componente de velocidad normal de partícula de un campo sonoro, que es proporcional al gradiente de P en la dirección normal a la superficie.

En la práctica, la integral de Kirchhoff-Helmholtz establece que en cualquier punto de escucha dentro del volumen V , la presión sonora $P(\mathbf{r},\omega)$ puede ser calculada si tanto la presión sonora como su gradiente son conocidos sobre la superficie que rodea el volumen. Esto puede utilizarse para sintetizar un campo sonoro dentro de la superficie S si se genera la distribución de presión apropiada $P(r_s,\omega)$. Este hecho es usado para la reproducción de sonido basada en WFS.

Si la superficie S degenera a un plano $z=z_1$, que separara el área de escucha del área de la fuente primaria, como se muestra en la figura 2.10 a), entonces la ecuación (2) se convierte en [13]:

$$P(r, \omega) = |z - z_1| \int_{S_1} P(r_s, \omega) \frac{1 + jk|r - r_s|}{2\pi|r - r_s|^3} e^{-jk|r - r_s|} dS_1 \quad (2.3)$$

Donde la geometría de esta ecuación se muestra en la figura 2.10 b), donde las superficies S_2 , S_3 y S_4 son completamente absorbentes. El campo acústico en el área de escucha puede ser generado por una distribución de fuentes secundarias en z_1 , cada fuente secundaria representa un dipolo, cuya señal de fuente está dada por la presión sonora primaria en esa localización.

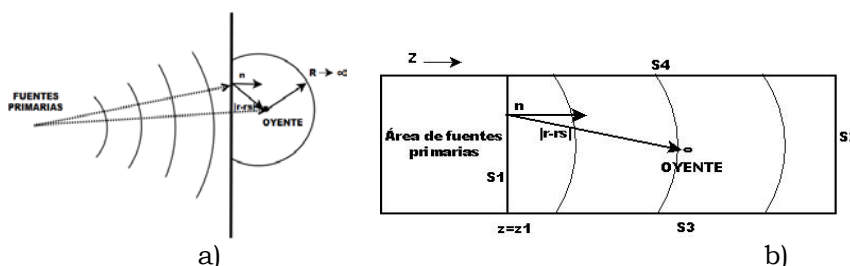


Figura 2.10 Geometría de la Integral de Kirchhoff-Helmholtz para superficie degenerada en un plano a); b) se muestra también la geometría del área de escucha con las superficies trasera y laterales completamente absorbentes.

Por tanto, es posible sintetizar físicamente los frentes de onda en cualquier posición de escucha al re-radiar la presión sonora, grabada por micrófonos en $z=z_1$, con altavoces que tengan características de dipolo, como se indica en la figura 2.11 a).

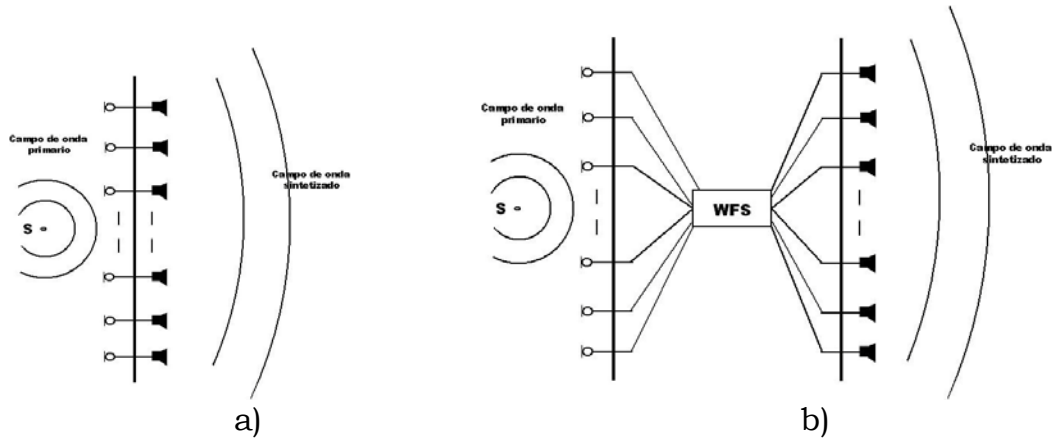


Figura 2.11 Diagrama de implementación práctica de WFS. a) Implementación de acuerdo a la ecuación 2.2 b) generalización de la técnica para implementación por extrapolación.

Un paso más allá es colocar los arreglos de transductores usados para grabación y sintetizar los frentes de onda en planos con diferentes coordenadas z_0 y z_1 , respectivamente, según se muestra en la figura 2.11 b. Entonces utilizando la ecuación (2.3), nuevamente las señales de los micrófonos deben transferirse a los altavoces a través de un procesador, simulando la propagación del frente de onda de z_0 y z_1 numéricamente. Este proceso se conoce como *extrapolación*. En esta configuración las posiciones de altavoces r_n en $z=z_1$, actúan como “posiciones de escucha” virtuales y entonces la señal enviada a cada altavoz en $z=z_1$. $P(r_n, \omega)$, es calculada procesando las señales de presión sonora $P(r_1, \omega)$ grabadas por los micrófonos en $z=z_0$, de acuerdo a la segunda integral de Rayleigh; ver más adelante, ecuación (2.9).

2.5.1.1. Primer esquema integral de Rayleigh

La primera integral de Rayleigh establece que el campo acústico en el espacio de escucha puede ser reconstruido a partir del campo sonoro original conociendo sólo la *velocidad* de partícula en el plano de medición y utilizando esas mediciones como fuentes de señal para una distribución de monopolos sobre el plano de reproducción [13] [37].

$$P(r, \omega) = \frac{jkpc}{2\pi} \iint_S u_n(r_s, \omega) \frac{e^{-jk|r-r_s|}}{|r-r_s|} dx dy \quad (2.4)$$

o en su forma discretizada:

$$P(r,\omega) = \frac{jk\rho c}{2\pi} \sum_n u_n(r_n, \omega) \frac{e^{-jk|r-r_n|}}{|r-r_n|} \Delta x \Delta y \quad (2.5)$$

Donde ω es la frecuencia angular, ρ es la densidad del medio homogéneo, c es la velocidad de propagación del sonido en el medio y el índice n indica los puntos de muestreo en el plano S . Las fuentes secundarias pueden estar construidas con altavoces monopulares pequeños con velocidad volumétrica:

$$U_n(\omega) = u_n(r_n, \omega) \Delta x \Delta y \quad (2.6)$$

Nótese que para un altavoz dinámico, arriba de su frecuencia de resonancia, la velocidad de volumen está relacionada al voltaje de entrada $E_n(\omega)$ en la bobina por:

$$U_n(\omega) = \frac{K_m}{j\omega} E_n(\omega) \quad (2.7)$$

donde K_m es una constante que depende de las propiedades electromecánicas de los sistemas de altavoces. A partir de las ecuaciones anteriores tenemos:

$$E_n(\omega) = \frac{j\omega}{K_m} \Delta x \Delta y \cdot u_n(r_n, \omega) \quad (2.8)$$

2.5.1.2. Segundo esquema integral de Rayleigh

La segunda integral de Rayleigh da relaciones similares entre la *presión* de sonido en el plano de medición y la distribución de dipolos en el plano de reproducción [13] [15]:

$$P(r, \omega) = \frac{jk}{2\pi} \iint_S P(r_s, \omega) \frac{1 + jk|r-r_s|}{jk|r-r_s|} \cos\phi \cdot \frac{e^{-jk|r-r_s|}}{|r-r_s|} dx dy \quad (2.9)$$

Con $\cos\phi = \frac{|z-z_s|}{|r-r_s|}$ donde $|z-z_s|$ es la distancia entre el plano de reproducción y el punto de observación. La forma discreta de esta ecuación está dada por:

$$P(r, \omega) = \frac{jk}{2\pi} \sum_n P(r_n, \omega) \frac{1 + jk|r-r_n|}{jk|r-r_n|} \cos\phi \frac{e^{-jk|r-r_n|}}{|r-r_n|} \Delta x \Delta y \quad (2.10)$$

En este caso las fuentes secundarias pueden ser construidas con pequeños altavoces sin baffle, y con una amplitud de volumen:

$$F_n(\omega) = P(r_n, \omega) \Delta x \Delta y$$

La relación entre la amplitud de volumen y el voltaje de entrada arriba de la frecuencia de resonancia mecánica está dada por:

$$F_n(\omega) = k_d E_n(\omega)$$

Donde K_d es una constante electromecánica del altavoz dipolo. De las ecuaciones anteriores tenemos:

$$E_n(\omega) = \frac{1}{K_d} P(r_n, \omega) \Delta x \Delta y \quad (2.11)$$

Nótese que con respecto al resultado de la primera integral de Rayleigh, ninguna ponderación en frecuencia es necesaria.

2.5.2. Obtención de las señales WFS

Para la derivación de la señal enviada a los altavoces, es considerada la geometría mostrada en la figura 2.12. El campo de presión de una fuente virtual, también conocido como *fente nocional* en r_m en el plano $z=z_0$ debe ser reconstruida en el plano horizontal del escucha, respecto a los oídos, utilizando un arreglo lineal de altavoces paralelo al eje x en el plano $z=z_1$. La línea que conecta la fuente y el escucha forma un ángulo θ respecto del eje z . Note que la fuente, el arreglo y el escucha están todos localizados en el plano $y=0$, esto es, a la misma altura. En la práctica el arreglo de altavoces estará colocado frecuentemente arriba de los niveles de la audiencia y la escena. En [13] se muestra que la altura del arreglo y_1 puede ser despreciada cuando esta altura es mucho más pequeña que las distancias entre la fuente y el arreglo y el escucha y el arreglo, lo que es frecuentemente el caso.

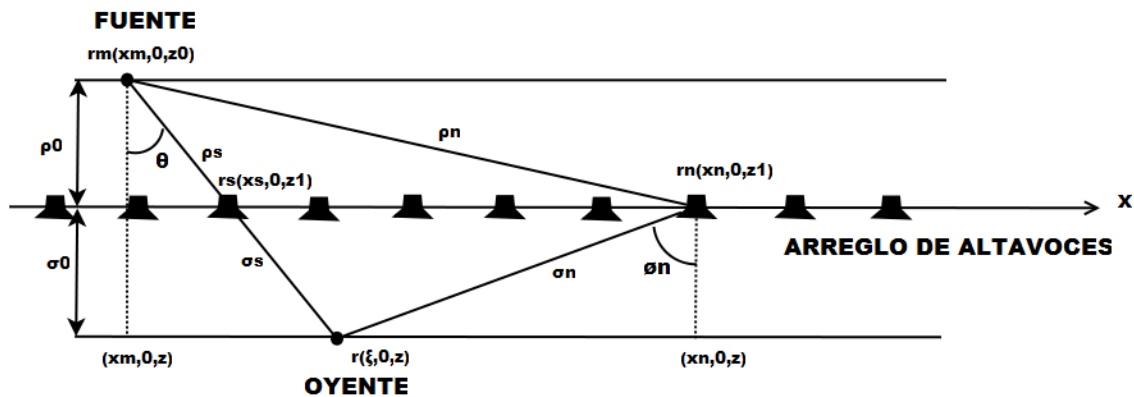


Figura 2.12 Configuración para WFS. El arreglo de altavoces en $z=z_1$ sintetiza el campo acústico de

una fuente en r_m para el plano de escucha para $z > z_1$.

La derivación dada aquí es una versión generalizada de la dada en el artículo antes mencionado [13] [21]. De acuerdo al teorema de Rayleigh, las señales enviadas al altavoz pueden ser escritas como una versión ponderada del campo de presión de la fuente nocional en la posición del arreglo.

$$Q(r_n, \omega) = A_n P(r_n, \omega) = A_n S(\omega) \frac{e^{-jk|r_n - r_m|}}{|r_n - r_m|} \quad (2.12)$$

Donde A_n es una función de ponderación que depende de la posición lateral del n-ésimo altavoz $A_n = A(x_n, \omega)$. Para sintetizar un frente de onda esférico, su formulación matemática debe ser considerada:

$$P(r, \omega) = S(\omega) \frac{e^{-jk|r - r_m|}}{|r - r_m|} \quad (2.13)$$

Donde $S(\omega)$ es el espectro de la fuente nocional. De acuerdo a la ecuación de Rayleigh, el frente de onda esférico puede ser sintetizado como:

$$P(r, \omega) = \sum_1^N \left[Q(r_n, \omega) G(\phi_n, \omega) \frac{e^{-jk|r - r_n|}}{|r - r_n|} \right] \Delta x \quad (2.14)$$

Donde N es el número de altavoces del arreglo, $Q(r_n, \omega)$ es la señal enviada al n-ésimo altavoz, ϕ_n es el ángulo entre el eje principal del n-ésimo altavoz y la línea que lo conecta con la posición del escucha, y Δx es la separación entre altavoces. Note que los únicos elementos desconocidos en el operador de síntesis son las señales enviadas de los altavoces. Utilizando la geometría de la figura anterior (2.12) la expresión anterior (2.14) puede reescribirse como:

$$\frac{e^{-jkr}}{r} = \sum_1^N \left[A_n G(\phi_n, \omega) \frac{e^{-jk|\rho_n + \sigma_n|}}{\rho_n \sigma_n} \right] \Delta x \quad (2.15)$$

Esta suma puede calcularse de manera aproximada utilizando el método de fase estacionaria. Físicamente esta aproximación significa que el frente de onda es sintetizado por todos los altavoces del arreglo juntos, pero una contribución dominante es dada por el altavoz posicionado en el punto de fase estacionaria; región desde la cual todas las fuentes contribuyen constructivamente con la misma fase. Después de una manipulación matemática considerable [13] [44], la señal enviada $Q(r_n, \omega)$ del n-ésimo altavoz puede encontrarse:

$$Q(r_n, \omega) = S(\omega) \frac{\cos \theta_n}{G(\theta_n, \omega)} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{|z - z_1|}{|z - z_0|}} \frac{e^{-jk|r_n - r_m|}}{\sqrt{|r_n - r_m|}} \quad (2.16)$$

Esta función contiene un factor de divergencia espacial cilíndrica $|r_n - r_m|^{-1/2}$ y entonces, la señal enviada del n-ésimo altavoz puede ser interpretada como una versión ponderada del campo de presión sonora en r_n , causado por una fuente lineal nocional en r_m . El hecho que sólo una línea horizontal sea utilizada en el plano de reconstrucción $z=z_1$ en el proceso de síntesis de frente de onda es “compensado” por la extensión espacial de la fuente nocional, de una fuente puntual a una fuente lineal vertical.

2.5.3. Potencial y limitaciones

2.5.3.1. Ventajas

La reproducción de audio a través de WFS ha ganado considerable atención desde que fue introducida primeramente por Berkhout [14] [15]. Una de las principales razones es el potencial que tiene para reproducir un campo sonoro con gran exactitud en cada lugar de la sala de escucha. Este no es el caso en las configuraciones de audio multicanal tradicionales como el sistema estéreo y los sistemas surround, las cuales no son capaces de generar la impresión espacial correcta más allá del lugar óptimo de la sala (sweet spot). Mediante WFS, este punto de escucha puede ser extendido a un área mucho más grande, a expensas de incrementar el número de altavoces [55] [17] [18].

Las capacidades teóricas de la WFS de crear un campo sonoro quasi-realista, o de recrear un sonido real existente, son grandes. Por ejemplo, es posible simular una cierta característica de directividad de la fuente virtual. Más aún, la localización de las fuentes secundarias (arreglo de altavoces) no tiene limitación para la creación de fuentes virtuales. WFS teóricamente permite la síntesis de fuentes virtuales tanto delante como detrás del arreglo. En particular, la creación de las llamadas fuentes enfocadas, las cuales están frente al arreglo, representa una diferencia significativa respecto a las técnicas de reproducción sonora convencionales. El diseño de la escena acústica está menos limitado por los problemas técnicos de reproducción en comparación con el estéreo. La simulación de una escena acústica real es más plausible. En la práctica esto significa que cuando el escucha está aproximándose a la localización de una fuente virtual, la amplitud se incrementa de una forma realista. En concordancia, la amplitud de una onda plana, la cual puede ser vista como una fuente a una distancia infinita, cambia menos en diferentes posiciones de escucha.

2.5.3.2. Limitaciones

Crear una copia de un campo sonoro no es completamente posible debido a limitaciones en la implementación. La aplicación práctica de WFS fuerza a que la fuente continua teórica, sea discretizada, lo que

causa una resolución espacial limitada. Como consecuencia, el campo sonoro es correctamente sintetizado sólo hasta una frecuencia límite, conocida como frecuencia “alias” (llamada la frecuencia espacial de alias). Este es uno de los principales problemas prácticos de la WFS, aunque no está restringido a este sistema de reproducción de audio en particular. La región espectral por arriba de esa frecuencia límite, está caracterizada por oscilaciones espurias de la amplitud en frecuencia y en el espacio. Esos patrones de interferencia, aunque no son pronunciados, pueden ser experimentados cuando la cabeza del oyente se mueve o camina en el área de escucha. Como la distribución de las fuentes secundarias para reproducción no es continua ni infinita, esos dos efectos limitan el comportamiento de un sistema WFS real [19].

- **Alias espacial.**

La distancia entre los transductores Δx (espaciados de modo que Δx es constante) define un muestreo en la frecuencia espacial para un campo de onda a nivel de grabación. Entonces el campo de onda reconstruido será físicamente correcto sólo hasta la frecuencia de Nyquist [13] [19] [21]:

$$f_{nyq} < \frac{c}{2\Delta x} \quad \text{o} \quad \Delta x < \frac{\lambda_{min}}{2} \quad (2.17)$$

donde λ_{min} es la longitud de onda más pequeña de interés y c es la velocidad del sonido. Arriba de la frecuencia de Nyquist, además del frente de onda correctamente reconstruido, cierta energía ondulatoria será emitida incorrectamente.

En la práctica, la recreación de campos de onda sin alias espacial es posible para frecuencias menores que la frecuencia de Nyquist mencionada (2.17). Esta frecuencia límite es determinada por la diferencia de tiempo entre las señales de dos altavoces sucesivos interfiriendo en la posición de escucha. Esta diferencia de tiempo depende del intervalo espacial de muestreo, esto es, de la separación entre micrófonos sucesivos, y entre altavoces sucesivos. Sin embargo, la máxima longitud de onda muestreada correctamente depende del máximo ángulo en el lado del micrófono. Por la misma razón, la máxima longitud que es correctamente recibida sin alias, también depende del ángulo máximo en el lado receptor. Esta frecuencia alias está dada de modo más general por [13] [56] [57]:

$$f_{al} = \frac{c}{2\Delta x | \text{sen} \alpha_{max} |} \quad (2.18)$$

Donde α_{max} indica el ángulo máximo entre el plano de la onda incidente y una superficie de arreglo de micrófonos. Si dicho ángulo es igual a 0 el frente de onda es perpendicular a la superficie del arreglo y el intervalo de muestreo espacial, Δx , puede ser visto como que tiende al infinito. En el

peor caso, cuando el ángulo es 90° , f_{al} es igual a la frecuencia de Nyquist. Afortunadamente, el sistema auditivo humano no es muy sensible para esos problemas de alias.

Cuando las fuentes virtuales sonoras son recreadas mediante WFS, el ángulo α_{max} puede ser dispuesto por diseño a un valor dado. La radiación de ondas planas en un ángulo más amplio que α_{max} es entonces suprimida, y el efecto de alias espacial puede ser evitado hasta la frecuencia f_{al} . La misma técnica puede ser aplicada para la grabación del campo de onda, donde micrófonos direccionales capturarán las ondas radiando hasta ese cierto ángulo.

- **Efectos de truncamiento.**

En teoría, la síntesis de los campos de onda viene de la suma de un número infinito de señales de altavoces. En la práctica, sin embargo, el arreglo de altavoces usado siempre tendrá una longitud finita. El arreglo finito puede ser visto como una ventana, a través de la cual la fuente virtual es visible, o invisible, al escucha. Por lo tanto existe un área que es iluminada por la fuente virtual, con una correspondiente área sombreada. Aplicando esta analogía, las ondas de difracción son originadas por los bordes del arreglo finito de altavoces. Esas contribuciones de error aparecen como post-ecos (retrasados en el tiempo) para las fuentes virtuales y pre-ecos (adelantados en el tiempo) para las fuentes enfocadas.

2.6. Resumen

A lo largo de este capítulo se han considerado dos asuntos importantes. Primero se ha descrito el sistema binaural de audición humano considerando los factores de los que depende la localización angular y de profundidad. Segundo, se ha hecho un recorrido a través de las diferentes técnicas de reproducción de sonido que existen, su historia, metas, logros y limitaciones. Se describe con mayor detalle el sistema de reproducción estereofónico convencional en cuanto a las condiciones de operación y los problemas derivados de su implementación práctica siendo el más importante el tener un área de escucha restringida (sweet spot). Se presenta el sistema de Síntesis de Frentes de Onda (WFS), que resuelve el problema del área de escucha restringida a costa del uso de un mayor número de altavoces y mayor complejidad de procesamiento.

3. Implementación del sistema

Ya se han descrito con anterioridad las características de los sistemas de reproducción de audio. El presente trabajo pretende lograr una mejora en los métodos de reproducción multicanal al volver independiente de la posición la reproducción sonora. Para ello se vale del uso de la teoría clásica del arreglo lineal de fuentes. La idea básica de esta implementación es lograr que mediante una serie de retardos se de una direccionalidad, variable mediante el software, al frente de onda reproducido por el arreglo lineal. Al aplicar esta idea a cada uno de los canales de una mezcla estereofónica y aplicarles respectivamente una direccionalidad izquierda y derecha se logra una equivalencia tal que la reproducción sea como la de dos fuentes ubicadas a cada lado del área de escucha. Sin embargo al ser un arreglo con retardos aplicados se logra una cierta independencia de la posición y consiguientemente una mejora respecto a la reproducción estereofónica normal.

3.1. Arreglo Lineal de altavoces

Sean N fuentes simples con separación d entre elementos adyacentes y longitud $L = (N-1)d$. La amplitud compleja de presión sonora que genera la fuente i -ésima, en un punto (r, θ) estaría dada por la siguiente expresión (ver figura 3.1):

$$P_i(r, t, \theta) = \frac{A}{r_i} e^{j(\omega t - kr_i)} \quad (3.1)$$

en la que el término r_i es la distancia de dicha fuente al punto, t es un tiempo dado y θ es el ángulo del punto a la normal del arreglo.

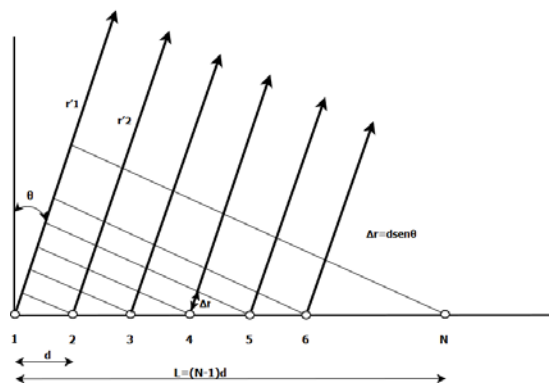


Figura 3.1 Geometría utilizada para obtener las características de radiación de un arreglo lineal de fuentes acústicas.

entonces la presión sonora total en el punto viene dada por la siguiente suma:

$$P_i(r, t, \theta) = \sum_1^N \frac{A}{r_i} e^{j(\omega t - kr_i)} \quad (3.2)$$

Haciendo ahora una aproximación para el campo lejano, cuando $r \gg L$, tenemos que las líneas de las fuentes al punto de radiación r_i son aproximadamente paralelas. En ese caso podemos considerar que:

$$r_i = r_1 - (i-1)\Delta r \quad (3.2a)$$

$$\Delta r = d \sin \theta \quad (3.2b)$$

y la distancia al centro del arreglo es:

$$r = r_1 - \frac{L\Delta r}{2d} \quad (3.2c)$$

en cuyo caso aproximamos cambiando r_i por r lo que deja:

$$P_i(r, t, \theta) = \sum_1^N \frac{A}{r} e^{j(\omega t - kr_i)} \quad (3.3)$$

Haciendo la sustitución de (3.2a) en (3.3) tenemos:

$$P_i(r, t, \theta) = \frac{A}{r} e^{j(\omega t - kr)} e^{j\frac{L\Delta r}{2d}} \sum_1^N e^{jk(i-1)\Delta r} \quad (3.4)$$

Donde $r = r_1 - \frac{1}{2}(L/d)\Delta r$ que es la distancia media al arreglo. Utilizando identidades trigonométricas tenemos:

$$P_i(r, t, \theta) = \frac{A}{r} e^{j(\omega t - kr)} \sum_1^N \frac{\text{sen}\left(\frac{N}{2}k\Delta r\right)}{\text{sen}\left(\frac{1}{2}k\Delta r\right)} \quad (3.5)$$

Que sobre $\theta=0$ implica:

$$P_i(r, t, \theta) = \frac{NA}{r} e^{j(\omega t - kr)} \quad (3.6)$$

Con factor direccional:

$$H(\theta) = \left| \frac{\text{sen}\left(\frac{N}{2}k\Delta r\right)}{N \cdot \text{sen}\left(\frac{1}{2}k\Delta r\right)} \right| \quad (3.7)$$

$H(\theta)$ es una función que tiene varios máximos que ocurren cuando el argumento de la función del denominador es múltiplo de π , entonces de acuerdo a (3.7) y (3.2a):

$$\text{sen}(\theta) = \left| \frac{2m\pi}{kd} \right| \quad \text{ó} \quad \Delta r = m\lambda \quad (3.7a)$$

La función dada por la ecuación 3.7 tiene además ceros cuando el argumento de la función del numerador haciendo esta consideración y de acuerdo a (3.2a):

$$\text{sen}(\theta) = \left| \frac{2n\pi}{kNd} \right| \quad \text{ó} \quad \Delta r = \frac{n\lambda}{N} \quad \text{con } m \neq \frac{n}{N} \quad (3.7b)$$

Los lóbulos menores (máximos secundarios) están, aproximadamente, en:

$$\text{sen}(\theta) = \frac{\left(n + \frac{1}{2} \right) \lambda}{Nd} \quad (3.8)$$

Cuando sólo se tiene un lóbulo mayor se requiere:

$$n=N-1, \text{ que implica que } \theta = \pi/2 \quad \text{y} \quad \frac{N}{N-1} = \frac{\lambda}{d}$$

Algunas aplicaciones requieren mover la direccionalidad del arreglo sin físicamente mover el mismo. Esto se puede lograr al aplicar un retardo electrónico en las señales que se alimentan a los altavoces. En ese caso tenemos que la ecuación para la presión debida a la fuente i -ésima (modificando 3.1) nos queda:

$$P_i(r, t, \theta) = \frac{A}{r_n} e^{j((\omega t + n\tau) - kr_n)} \quad (3.9)$$

Con el factor direccional:

$$H(\theta) = \frac{\left| \text{sen} \left(\frac{Nkd}{2} \left(\text{sen}\theta - \frac{c\tau}{d} \right) \right) \right|}{\left| N \cdot \text{sen} \left(\frac{kd}{2} \left(\text{sen}\theta - \frac{c\tau}{d} \right) \right) \right|} \quad (3.10)$$

donde ahora el lóbulo mayor apunta a:

$$\text{sen}(\theta_0) = \frac{c\tau}{d} \quad (3.11)$$

Entonces cuando el ángulo es $\pi/2$ (máxima lateralidad del arreglo) tenemos que:

$$\tau = \frac{d}{c}$$

Con respecto a los programas implementados en esta tesis y descritos en la sección 3.4, tenemos los siguientes datos:

$$f_s = 44100 \text{ Hz} \quad \text{por lo que} \quad T_s = f_s^{-1}$$

Entonces el retardo aplicado a cada altavoz subsecuente del arreglo durante el procesamiento de los programas es:

$$\tau_p = nT_s$$

Donde n es el número de muestras que el programa retarda una señal dada. Ahora la ecuación 3.11 queda:

$$\text{sen}(\theta_0) = \frac{cnT_s}{d} \quad (3.12)$$

En la siguiente tabla vemos la relación entre el número de muestras de programa, el retardo en tiempo y el ángulo de directividad del arreglo. Dado que en general se obtienen números de muestras fraccionales, se redondeó al entero más próximo; se muestran los ángulos reales de direccionalidad bajo esta corrección. Los datos son conforme a la implementación experimental ($d = 0.35 \text{ m}$, $c = 343 \text{ m/s}$).

θ [°]	d [m]	n [muestras]	$n_{\text{redondeadas}}$ [muestras]	T_p [us]	θ_{real} [°]
10	0.35	7.8	8	0.06	10.2
20	0.35	15.3	15	0.34	19.4
30	0.35	22.5	23	0.52	30.7
40	0.35	28.9	29	0.66	40.1
50	0.35	34.4	34	0.77	49.0
60	0.35	38.9	39	0.88	60.1
70	0.35	42.2	42	0.95	68.9
80	0.35	44.3	44	1.00	77.9
90	0.35	45	45	1.02	90

Tabla 3.1 Relación entre el tiempo de retardo aplicado a los altavoces del arreglo lineal, el número de muestras de los programas de procesamiento y el ángulo de direccionalidad del arreglo. Se muestra el ángulo real de direccionalidad respecto al número de muestras redondeado.

3.2. Simulaciones

Se realizaron simulaciones con el objetivo de obtener una idea, intuitiva, sobre el modo en que habría de desempeñarse la implementación propuesta para la reproducción de señales sonoras mediante un arreglo de altavoces. Se simuló el comportamiento de los frentes de onda emitidos por ocho fuentes monopares, en una configuración de arreglo lineal, en analogía al arreglo implementado experimentalmente.

En las simulaciones se tuvieron en cuenta las siguientes consideraciones:

- Se tienen 8 fuentes monopolares que radian una señal senoidal cuya frecuencia es de 900 Hz.
- La separación entre las fuentes es de 0.54 m.
- La señal entre altavoces adyacentes tiene un tiempo de retardo variable que corresponde a ángulos de direccionalidad de 0, 30, 60 y 90 grados.

El programa fue elaborado en Matlab (frentdonda.m ver apéndice), basándose en la ecuación 3.9, de modo que se tiene una línea de retardos variable que confiere al arreglo una direccionalidad también variable como función del retardo. Los resultados se aprecian en las figuras 3.2 a 3.5, donde el área de observación horizontal abarca 9 metros, tanto a la izquierda, como a la derecha del centro del arreglo, aún cuando esto es mucho mayor que el área de escucha considerada. En el eje vertical se consideran 60 longitudes de onda.

Para el arreglo sin retardos, figura 3.2, puede observarse que se tiene esencialmente un frente de onda plano en la región de observación. La figura corresponde a las curvas de nivel, en unidades lineales, para un instante de tiempo del campo acústico de presión sonora. Para las figuras 3.3 a 3.5 que corresponden a diferentes, y progresivos retardos aplicados (en pasos de 30°), puede notarse que el frente de onda conserva esencialmente su geometría plana y que adquiere diferentes grados de direccionalidad con respecto a la normal del arreglo (ubicado en la parte inferior horizontal de las figuras). También puede observarse la aparición de lóbulos laterales de gran amplitud.

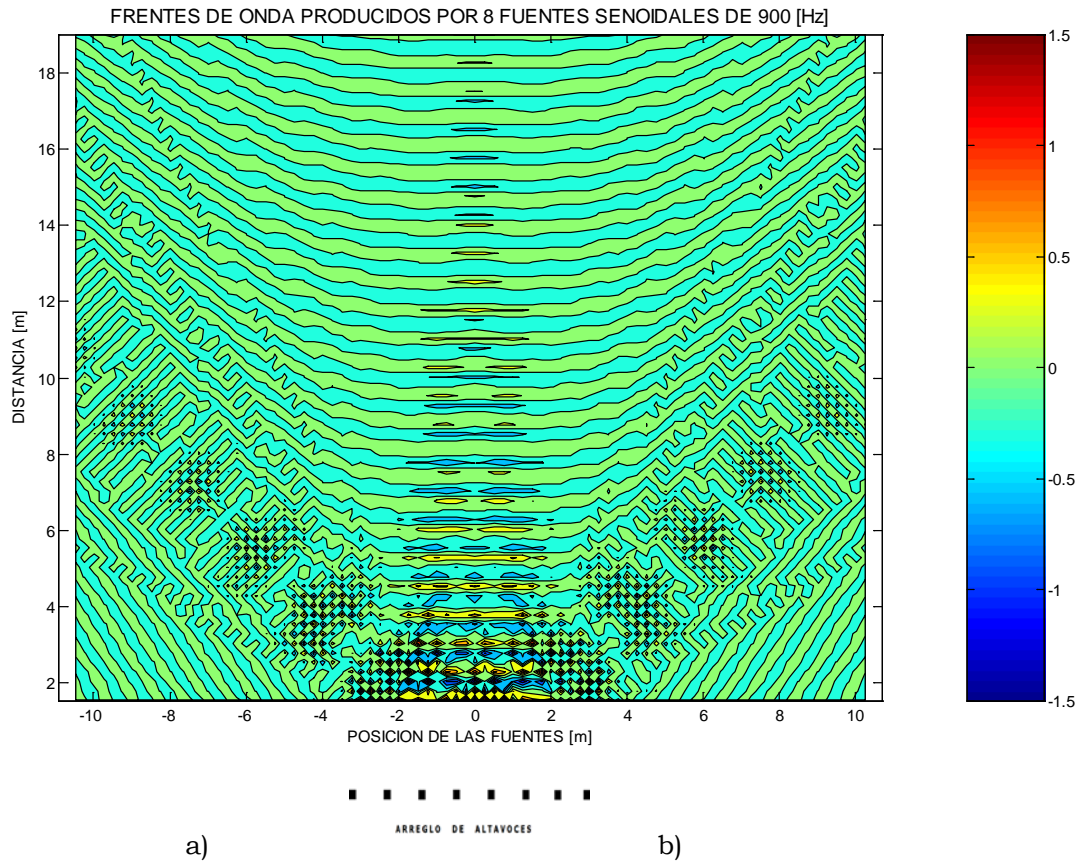


Figura 3.2 Representación en curvas de nivel para el campo de presión sonora instantánea producida por un arreglo de 8 fuentes monopulares en arreglo lineal, con separación de 0.54 m, con un retardo correspondiente a una direccionalidad de 0° .

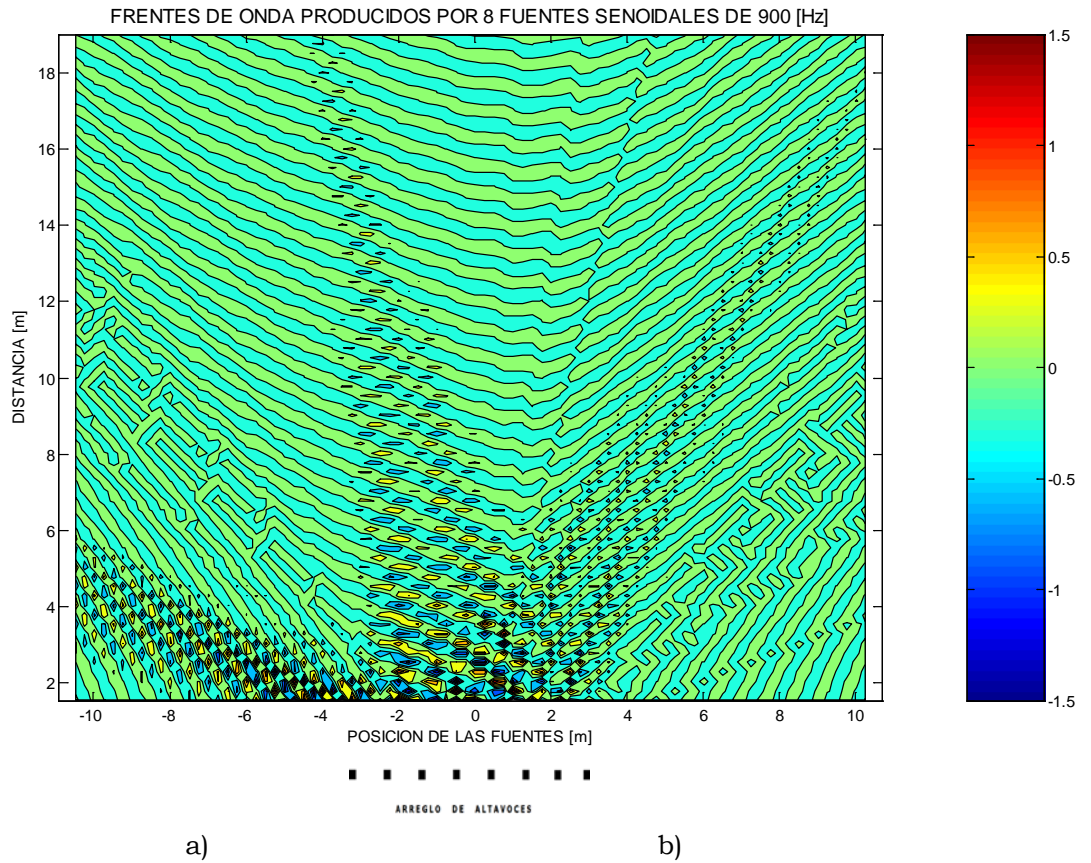


Figura 3.3 Representación en curvas de nivel para el campo de presión sonora instantánea producida por un arreglo de 8 fuentes monopulares en arreglo lineal, con separación de 0.54 m, con un retardo correspondiente a una direccionalidad de 30°.

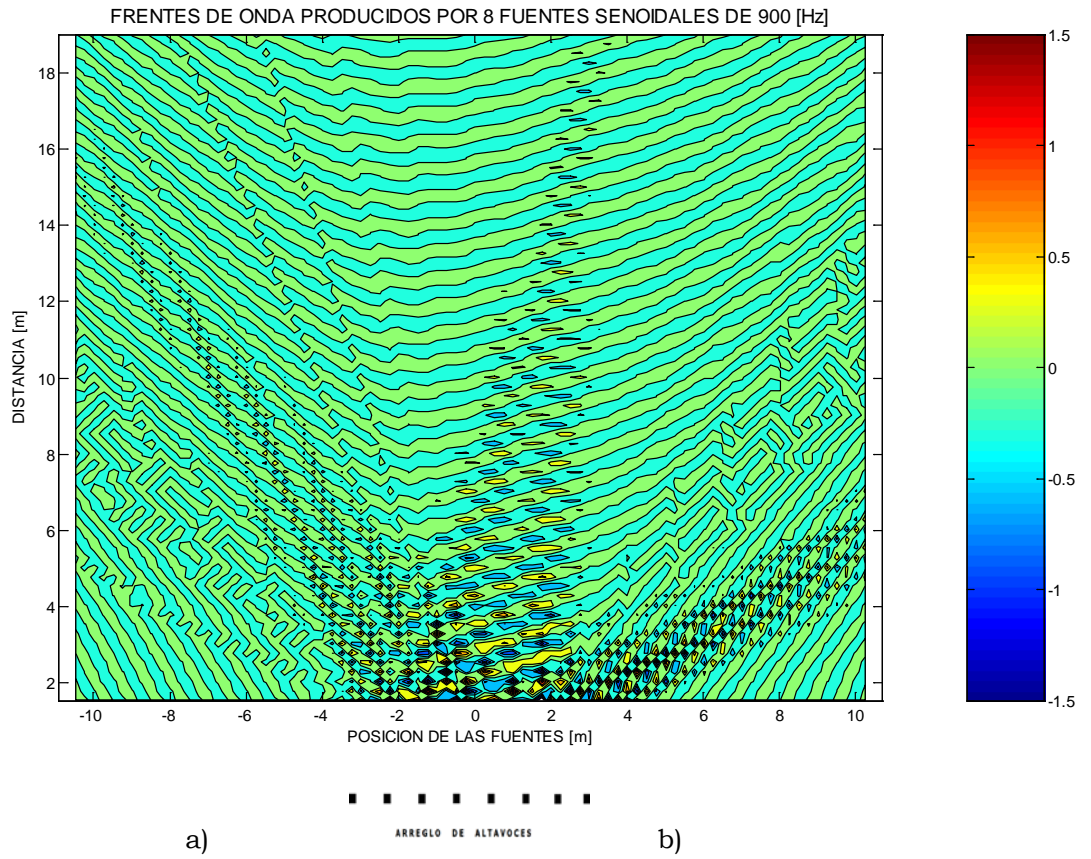


Figura 3.4 Representación en curvas de nivel para el campo de presión sonora instantánea producida por un arreglo de 8 fuentes monopulares en arreglo lineal, con separación de 0.54 m, con un retardo correspondiente a una direccionalidad de 60° .

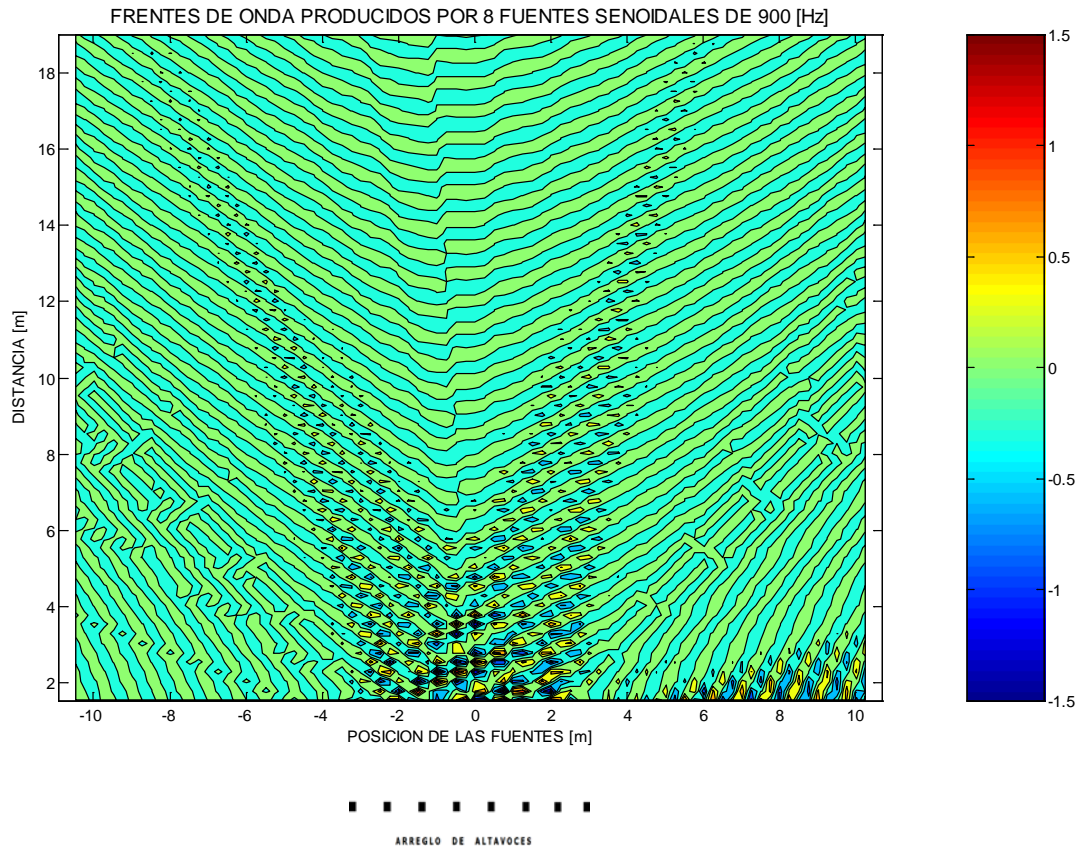


Figura 3.5 Representación en curvas de nivel para el campo de presión sonora instantánea producida por un arreglo de 8 fuentes monopulares en arreglo lineal, con separación de 0.54 m, con un retardo correspondiente a una direccionalidad de 90° .

3.3. Software e interfaz de audio

En este trabajo se desarrollaron programas para la reproducción de sonido multicanal por medio de un arreglo de 8 altavoces, controlados por una interfaz de audio multicanal y una computadora personal realizando el procesamiento de las señales en tiempo real (ver figura 3.6) [58] [59]. La programación se realizó en lenguaje C, utilizando varios modelos de computadora Apple Macintosh (Mac-mini Power PC con 1GB de RAM y MAC 3.06 GHz Intel Core 2 Duo, 4 GB de RAM), sistema operativo MacOS X (10.4 y 10.5). Sin embargo, la programación es multiplataforma y fácilmente transportable a sistemas Windows, Linux y otros, debido a que se utilizaron las siguientes herramientas abiertas de programación: compilador GNU gcc, bibliotecas fftw3 (fftw3 3.2.2 <http://www.fftw.org>) y portaudio V18 (<http://www.portaudio.com>). Para la reproducción multicanal en 8 altavoces se utilizó la interfaz de audio de 8 canales de entrada y 8 de salida MOTU 896-HD (<http://www.motu.com/audio>); aunque también es posible utilizar muchas otras interfaces de audio multicanal soportadas por la biblioteca portaudio.

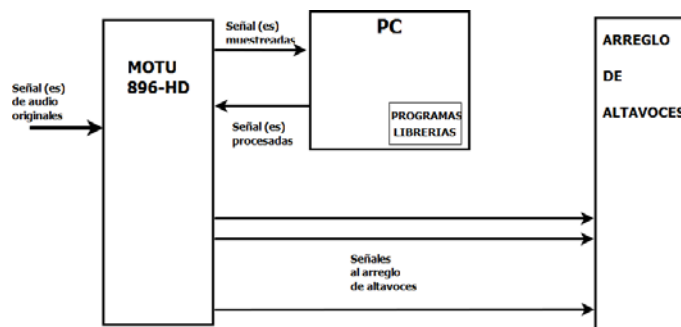


Figura 3.6 Diagrama general del sistema implementado. Se muestra la interacción del hardware, software y las señales procesadas.

3.4. Programas implementados

Para los propósitos de la reproducción de señales de audio, en tiempo real, se implementó una serie de programas que son tratados en esta sección. El procesamiento de las señales se realiza en su mayor parte en el dominio de la frecuencia, aunque algunos programas trabajan sobre las señales en el dominio del tiempo. Los programas desarrollados permiten realizar los siguientes tipos de procesamiento:

1. Reproducción monofónica en arreglo circular, con señal secuenciada individualmente entre los altavoces.
2. Reproducción monofónica en arreglo lineal, con

direccionalidad izquierda, derecha o mixta.

3. Reproducción estereofónica en arreglo lineal, con direccionalidad izquierda+derecha o arbitraria.

El primer programa representa un caso especial, ya que se activa un solo altavoz a la vez, y por lo tanto su funcionamiento correcto se comprueba directamente, sin necesidad de realizar mediciones acústicas. Para los fines de la evaluación experimental del desempeño de los demás programas, es necesario probar únicamente uno de ellos. La razón de esto obedece a que todos utilizan el mismo principio acústico de reproducción en arreglo lineal.

3.4.1. circulación.c

Este programa hace circular una señal de entrada en un arreglo de altavoces, que pueden estar acomodados en arreglo lineal o circular de manera que únicamente un altavoz reproduce la señal en un instante de tiempo. Para ello se hace una equipartición del bloque de muestras de señal de audio tomadas de modo que a cada altavoz le toca reproducir una parte del bloque. Esto se controla mediante una variable ('retardo', ver programa en apéndice) que depende proporcionalmente del tamaño del bloque muestreado y con dependencia inversamente proporcional al número de canales de salida y entrada. Como límite superior está en relación con el tamaño del bloque de muestras y como límite inferior el periodo de muestreo más una latencia inherente al procesamiento mismo. Su funcionamiento es de acuerdo al siguiente esquema (figura 3.7).

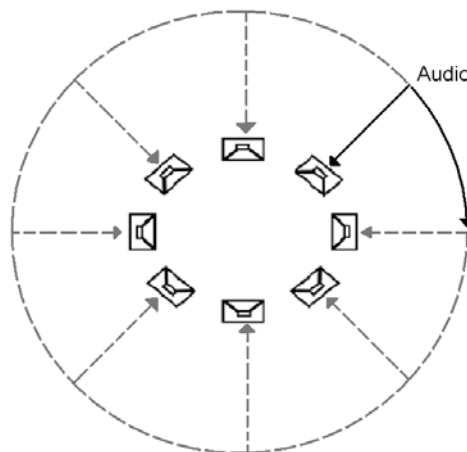


Figura 3.7 Esquema del programa 'circulacion.c'. El programa hace circular una señal de audio en un arreglo de altavoces, que pueden estar acomodados según la figura o linealmente, etc.

3.4.2. retardosizq.c

Este programa implementa una línea de retardos en un arreglo lineal de altavoces sobre una señal de entrada monofónica. El retardo se aplica de derecha a izquierda del arreglo. Así pues se proporciona al arreglo una lateralidad, de ángulo variable, en función del tiempo de retardo aplicado a cada altavoz sucesivo, hacia la izquierda. Para un ángulo de reproducción deseado debe aplicarse un retardo correspondiente, que el usuario define, de acuerdo a la tabla 3.1 para las mediciones hechas y de acuerdo a la ecuación 3.12 para otras condiciones de reproducción, teniendo como límite superior el tamaño del bloque muestreado entre el número de canales de salida y como límite inferior una muestra del bloque más una latencia propia del procesamiento en los dispositivos, según se muestra en el siguiente esquema (figura 3.8).

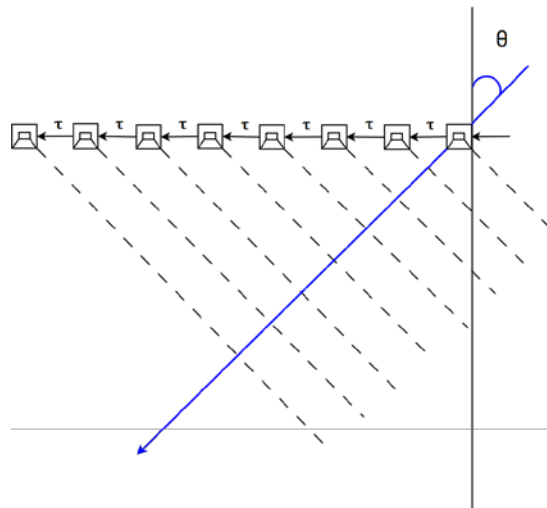


Figura 3.8 Esquema del programa 'retardosizq.c'. El programa aplica un retardo τ de duración variable a la señal alimentada a cada altavoz sucesivo, de modo que se obtiene un efecto de lateralidad de ángulo variable.

3.4.3. retardosder.c

Este programa implementa una línea de retardos en un arreglo lineal de altavoces sobre una señal de entrada monofónica. El retardo se aplica de izquierda a derecha. De esta manera se proporciona al arreglo una lateralidad, de ángulo variable. El ángulo de reproducción es función del tiempo de retardo aplicado a cada altavoz sucesivo, hacia la derecha. Para un ángulo de reproducción deseado debe aplicarse un retardo correspondiente, de acuerdo a la tabla 3.1 para las mediciones hechas y de acuerdo a la ecuación 3.12 para otras condiciones de reproducción,

aplicando los mismos límites que en el programa `retardosizq.c`, según se muestra en el siguiente esquema (figura 3.9).

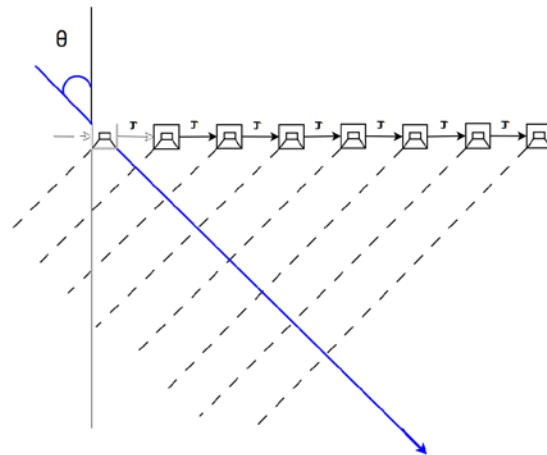


Figura 3.9 Esquema del programa 'retardosder.c'. El programa aplica un retardo τ de duración variable a la señal alimentada a cada altavoz sucesivo, de modo que se obtiene un efecto de lateralidad de ángulo variable con direccionalidad derecha.

3.4.4. `retizqder.c`

Este programa, mediante la aplicación de un retardo variable, implementa una línea de retardos en un arreglo lineal de altavoces sobre una señal de entrada monofónica. El retardo puede ser aplicado al arreglo de derecha a izquierda o viceversa. De esta manera se puede proporcionar una lateralidad hacia la izquierda o derecha del arreglo, controlando desde el software. Para un ángulo de reproducción deseado debe aplicarse un retardo correspondiente, de acuerdo a la ecuación 3.12 para otras condiciones de reproducción. En este caso el retardo puede ser negativo o positivo, correspondiendo a un retardo de cero muestras la reproducción frontal con 0° de direccionalidad, las limitaciones en los retardos son la mitad que en los dos programas anteriores dado que el retardo puede ser negativo o positivo (a la izquierda o derecha).

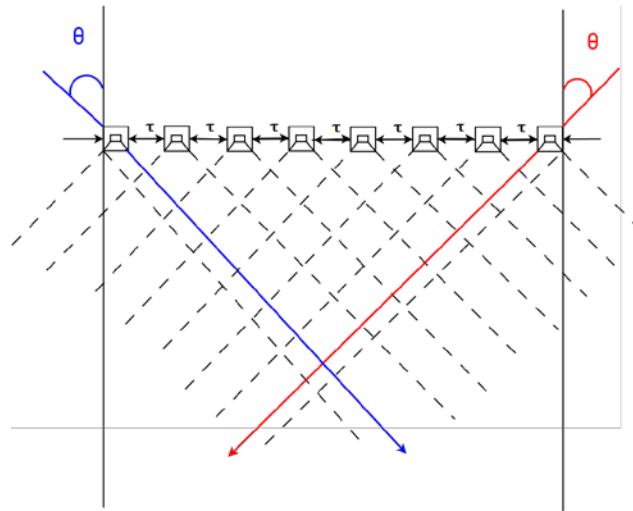


Figura 3.10 Esquema del programa 'retizqder.c'. El programa aplica un retardo τ de duración variable a la señal alimentada a cada altavoz sucesivo. Puede ser aplicado hacia cualquier lado del arreglo de modo que se obtiene un efecto de lateralidad de ángulo variable con direccionalidad derecha o izquierda.

3.4.5. estereo.c

Este programa implementa una línea de retardos en un arreglo de altavoces sobre una señal de entrada estereofónica (dos canales). Cada señal de entrada se alimenta al arreglo en una dirección (e.g. canal izquierdo de izquierda a derecha; canal derecho viceversa). De esta manera, se proporciona al arreglo una lateralidad hacia la izquierda para un canal y a la derecha para el otro canal, según se muestra. Con este programa cada canal puede ser movido solo hacia un lado del arreglo. El ángulo de reproducción es función del tiempo de retardo aplicado a cada altavoz sucesivo, hacia la derecha. Para un ángulo de reproducción deseado debe aplicarse un retardo correspondiente, de acuerdo a la tabla 3.1 para las mediciones hechas y de acuerdo a la ecuación 3.12 para otras condiciones de reproducción, aplicando los mismos límites que en el programa retardosizq.c

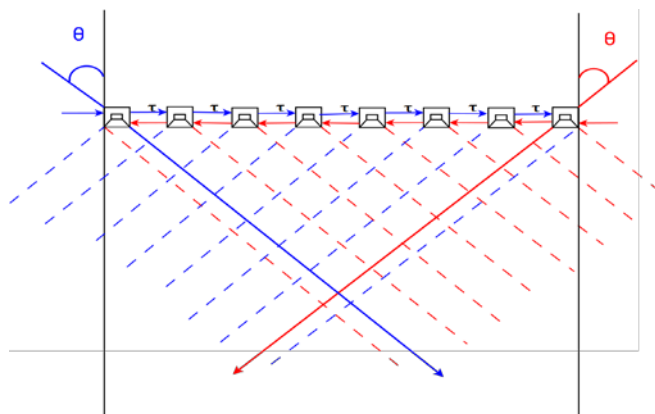


Figura 3.11 Esquema del programa 'estereo.c'. El programa aplica un retardo τ de duración variable a una señal alimentada a cada altavoz sucesivo. Cada canal de la señal de audio se aplica en una dirección del arreglo. Se obtiene de este modo lateralidad para cada canal de la señal estereofónica.

De este modo, al dar lateralidad a cada canal de la señal estereofónica podemos mover la ubicación virtual de la fuente para cada uno de ellos. Esto es análogo a modificar la separación de dos altavoces reproduciendo una señal estereofónica, manteniendo la restricción de que cada fuente está a un lado del área de audición.

3.4.6. estereom.c

Este programa implementa una línea de retardos en un arreglo de altavoces sobre una señal de entrada estereofónica (dos canales). Cada señal de entrada se alimenta al arreglo en una dirección (e.g. canal izquierdo de izquierda a derecha; canal derecho viceversa). De manera que se puede proporcionar al arreglo una lateralidad arbitraria para cada canal. Esto es, ambos canales pueden moverse en direcciones opuestas del arreglo o ambos hacia un mismo lado del mismo.

Análogamente al programa descrito en el punto anterior, puede efectuarse, mediante modificar el ángulo de lateralidad del arreglo, una modificación de la ubicación virtual de las fuentes. Sin embargo en este caso no se restringe a que cada una de ellas esté de un lado del área de audición, sino que ambas pueden estar del mismo lado, modificando con ello el área misma de audición. Para un ángulo de reproducción deseado debe aplicarse un retardo correspondiente, de acuerdo a la ecuación 3.12, un retardo de cero muestras corresponde a una reproducción frontal con direccionalidad de 0° . Nuevamente aquí el retardo puede ser positivo o negativo, teniéndose como límites los del programa retizqder.c.

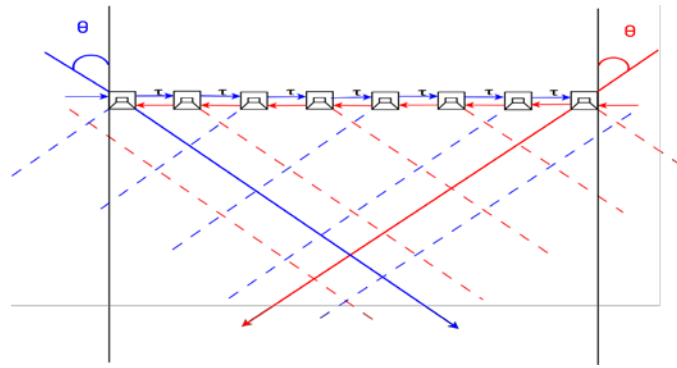


Figura 3.12 Esquema del programa 'estereom.c'. El programa aplica un retardo τ de duración variable a la señal alimentada a cada altavoz sucesivo. Cada canal de la señal de audio se aplica en una dirección del arreglo. Se obtiene de este modo lateralidad arbitraria para cada canal de la señal estereofónica.

3.5. Resumen

Se ha presentado en este capítulo la idea general del sistema propuesto en esta tesis. Se presentan simulaciones para la radiación de un arreglo de altavoces análogo al utilizado. Se ha hecho un repaso de la teoría del arreglo lineal de fuentes acústicas y se ha considerado el caso de un arreglo con retardos en la señal aplicada a cada fuente. Se ha considerado el equipo utilizado para la implementación de esta propuesta. Finalmente se han presentado los programas implementados, de los cuales se han explicado su funcionamiento y limitaciones.

4.Resultados

4.1. Pruebas objetivas y mediciones

Las mediciones se realizaron en condiciones de campo acústico libre, en la Cámara Anecoica del Laboratorio de Acústica y Vibraciones del CCADET-UNAM. La Figura 4.1 muestra un esquema del arreglo de medición. Se tiene un arreglo lineal de 8 altavoces auto-amplificados Event TR5 (con una separación de 0.54 m entre ellos), que reproducen una señal de entrada de audio monofónica, a la que se procesa mediante una línea de retardos, utilizando el equipo mencionado anteriormente, para que se tenga una direccionalidad por la derecha en el sonido reproducido.

Simultáneamente, se realizaron mediciones con un arreglo de 8 micrófonos (con una separación de 0.33 m entre ellos), conectados a una interfaz de audio MOTU 8-Pre y ésta a una computadora Apple Macintosh (mencionada antes), ejecutando un programa de medición de respuesta en frecuencia multicanal, con una salida y ocho entradas. Este programa fue desarrollado también en lenguaje C, con las herramientas de programación descritas anteriormente.

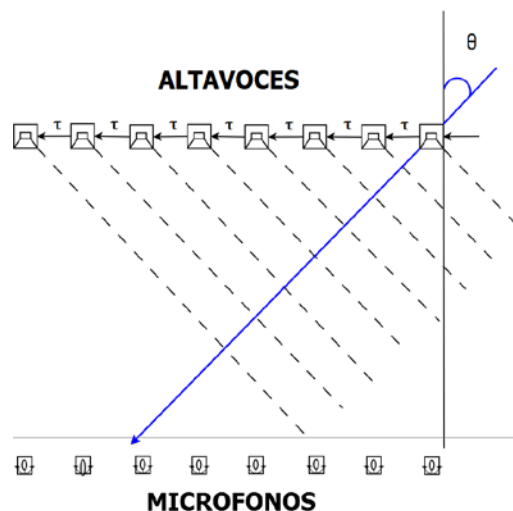


Figura 4.1 Arreglo lineal de altavoces y arreglo lineal de micrófonos, dispuestos para la evaluación experimental de la reproducción de una señal de audio monofónica con diferentes ángulos de incidencia aparente.

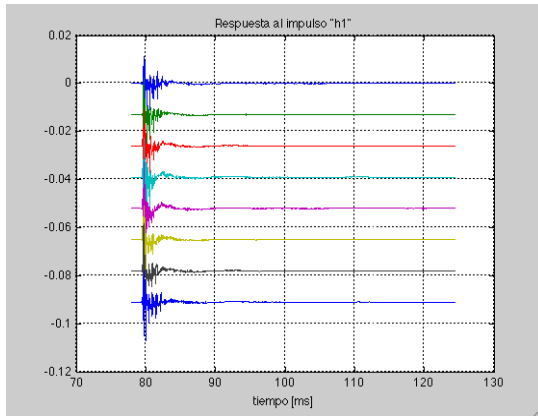
El programa produce de manera repetida una señal de prueba, que puede ser un barrido en frecuencia (chirp) o bien ruido pseudo-aleatorio. Esta señal se alimenta al sistema de reproducción multicanal de 8 altavoces y el sonido captado por los 8 micrófonos se utiliza para calcular los estimadores H_1 y H_2 de la respuesta en frecuencia (que son las relaciones del espectro cruzado de las señales de entrada y salida entre el autoespectro de la señal de entrada y del autoespectro de la señal de salida entre el espectro cruzado de las señales de salida y entrada respectivamente [ecuaciones 4.1]) entre la señal de entrada (señal de prueba) y la señal captada por cada uno de los 8 micrófonos. El programa guarda los datos obtenidos de la medición en formato half-complex. Los datos son analizados mediante un programa en Matlab que calcula estas funciones.

$$H_1 = \frac{G_{xy}}{G_{xx}} \qquad H_2 = \frac{G_{yy}}{G_{yx}} \qquad (4.1)$$

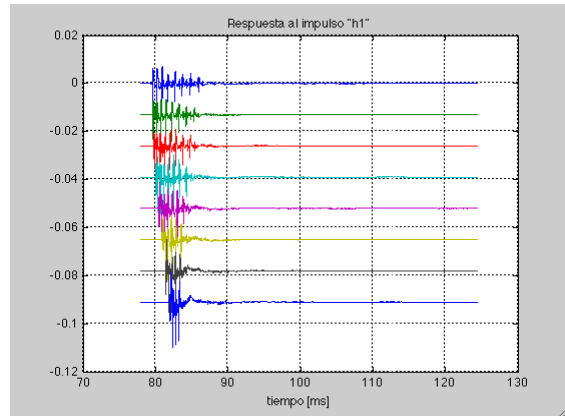
Se realizaron mediciones de este tipo con 10 ángulos de reproducción diferentes, de 0° a 90°, en intervalos de 10°. Las Figuras 4.2, 4.3 y 4.4 muestran las respuestas a impulso obtenidas por transformación inversa de Fourier, de las funciones definidas en 4.1 [42], así como la magnitud y fase de las respuestas en frecuencia correspondientes; esto, para ángulos de reproducción de 0°, 30°, 60° y 90°.

En la Figura 4.2 se observan retrasos que aumentan progresivamente conforme aumenta el ángulo de incidencia, de acuerdo a lo esperado; así como una mayor separación temporal de los impulsos que corresponden a los 8 altavoces, para los micrófonos más cercanos al primer altavoz del arreglo. Esto último indica limitaciones de desempeño para ángulos grandes.

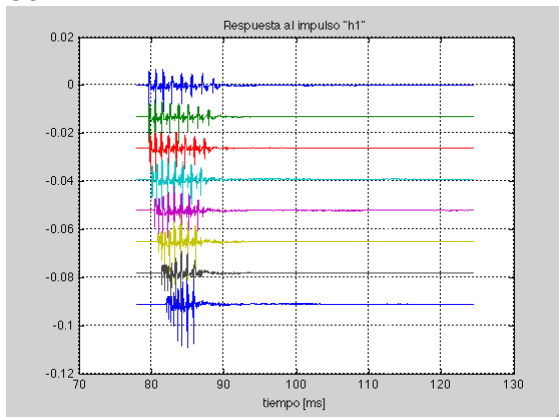
0°



30°



60°



90°

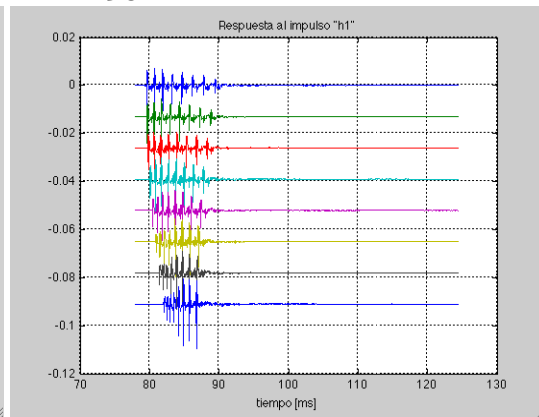
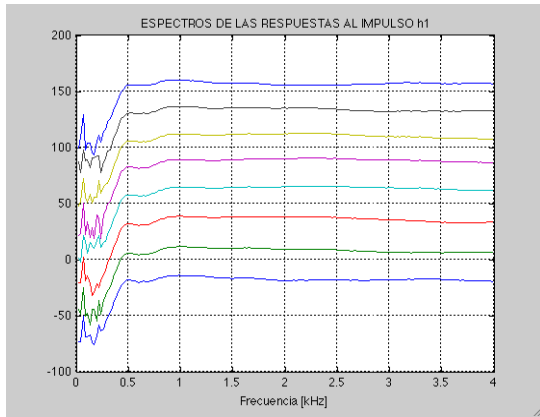


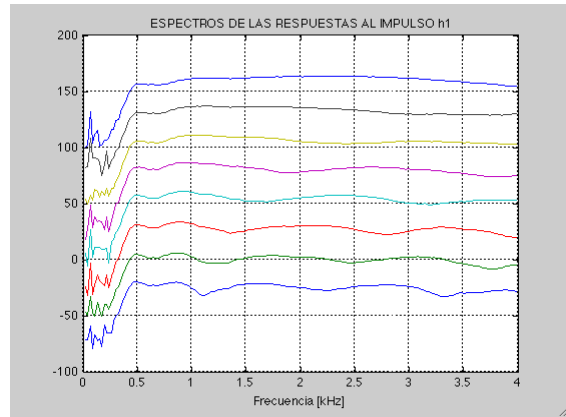
Figura 4.2. Respuestas a impulso entre la señal de entrada de audio monofónica y un arreglo lineal de 8 micrófonos en condiciones de campo libre (cámara anecoica), para diferentes ángulos de reproducción.

La Figura 4.3 muestra una reducción de respuesta por debajo de 500 Hz, asociada a la respuesta en bajas frecuencias de una fuente acústica monopolar; la respuesta se hace plana arriba de los 500 Hz, con efecto de filtro peine para los ángulos de incidencia mayores. La Figura 4.4 muestra mayores diferencias de fase para ángulos de incidencia mayores, conforme a lo esperado; pero con respuesta irregular en el caso de 90°, que indica también límites para ángulos grandes.

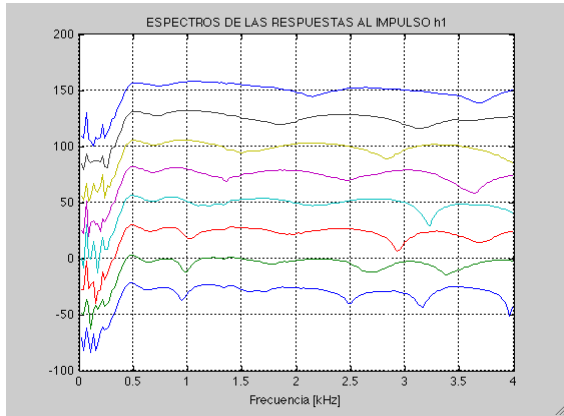
0°



30°



60°



90°

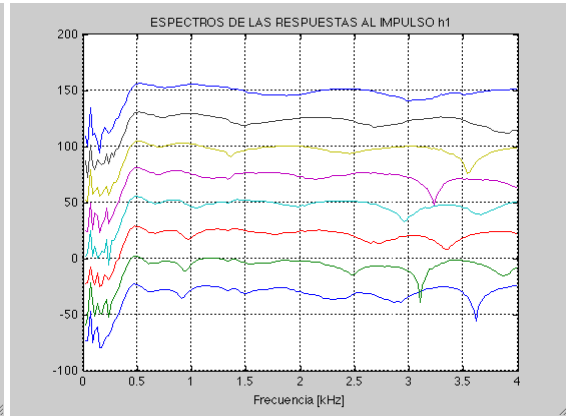
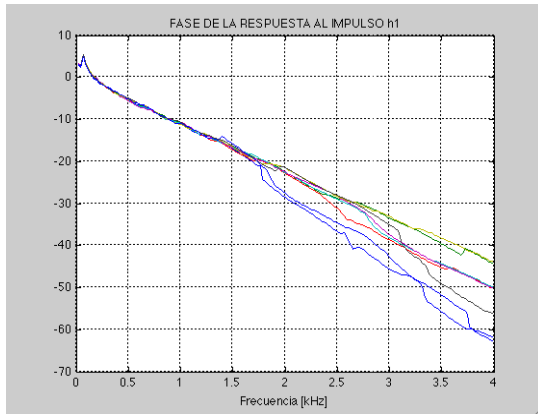
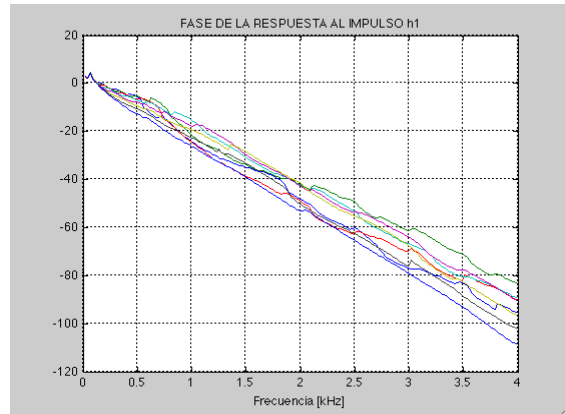


Figura 4.3. Magnitud de las respuestas en frecuencia entre la señal de entrada de audio monofónica y un arreglo lineal de 8 micrófonos en condiciones de campo libre (cámara anecoica), para diferentes ángulos de reproducción.

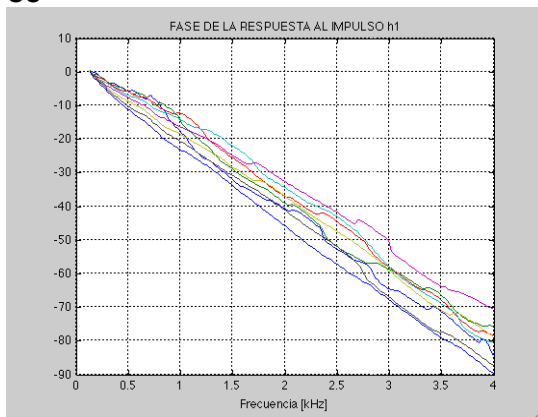
0°



30°



60°



90°

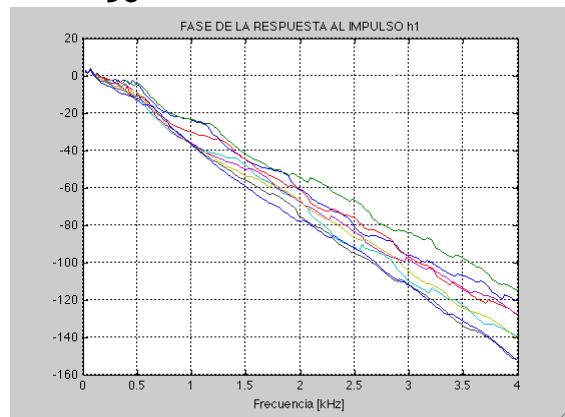


Figura 4.4. Fase de las respuestas en frecuencia entre la señal de entrada de audio monofónica y un arreglo lineal de 8 micrófonos en condiciones de campo libre (cámara anecoica), para diferentes ángulos de reproducción.

También se calculó el retraso de grupo, para las 8 posiciones de micrófono y para los 10 ángulos de reproducción. Esto se realizó a partir de la regresión lineal de mínimo error cuadrático de la respuesta de fase $\varphi(\omega)$, y considerando la definición del retardo de grupo $\tau_g = -d\varphi/d\omega$, que indica tomar el valor negativo de la pendiente lineal de la respuesta de fase en función de la frecuencia [43]. Los resultados se muestran en la Figura 4.5. Para ángulos de incidencia de hasta 60°, aproximadamente; se observa un comportamiento cercano al esperado. Esto en todas las posiciones de micrófono, lo cual demuestra que la zona de audición es amplia, según lo deseado. Para ángulos mayores de 60°, el desempeño es pobre en general.

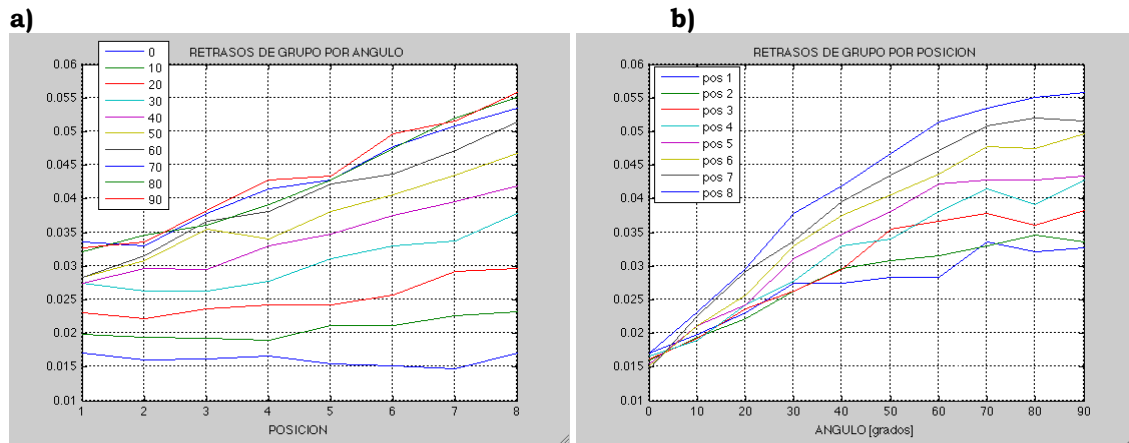


Figura 4.5. Retraso de grupo entre la señal de entrada de audio monofónica y un arreglo lineal de 8 micrófonos en condiciones de campo libre (cámara anecoica): a) en función de la posición de cada micrófono para diferentes ángulos de reproducción; b) en función del ángulo de reproducción para diferentes posiciones de micrófono. La escala vertical tiene unidades de tiempo, en segundos.

5. Conclusiones

- A partir de las mediciones de la reproducción direccional del arreglo de altavoces, se puede ver que los retrasos aumentan progresivamente de manera proporcional con el ángulo de incidencia, lo cual confirma que efectivamente se ha dado una direccionalidad al sonido emitido por el arreglo. Sin embargo, para ángulos de direccionalidad grandes se encuentran limitaciones dadas por la mayor separación temporal de los impulsos medidos en las posiciones próximas al primer altavoz del arreglo.
- La respuesta en frecuencia muestra que el arreglo se comporta pobremente debajo de 500 Hz, en general (lo que da una frecuencia límite inferior de operación). Esto sin embargo, no es un problema,
- dado que las bajas frecuencias pueden ser reproducidas en un canal separado debido a sus características direccionales. Para ángulos de incidencia grandes, se tienen efectos de filtro peine, lo que limita también el desempeño del arreglo en altas frecuencias, en general se tiene un buen comportamiento, para los diferentes ángulos en el rango de 500 a 3000 Hz.
- A partir de las mediciones de diferencias de fase y de retrasos de grupo, se confirma la existencia de limitaciones en la correcta operación del arreglo, para las señales reproducidas con ángulos arriba de 60° , lo cual establece un ángulo límite de direccionalidad aplicable.
- Los retrasos de grupo medidos son conforme a lo que se esperaba, y muestran que la zona de audición puede ser aumentada mediante este tipo de procesamiento, en comparación con los sistemas convencionales de reproducción estereofónica.

5.1. Perspectivas

- El sistema implementado en esta tesis puede emplearse para aplicaciones de reproducción de mezclas estereofónicas de audio de alta calidad, como pueden ser musicales o aplicaciones multimedia que requieran una amplia zona de audición, en beneficio de una audiencia numerosa.
- La aplicación de este tipo de procesamiento, por ejemplo, a la parte frontal de mezclas surround 5.1, que reforzaría la calidad de

reproducción y la haría más confortable en salas cinematográficas, es viable. Existen perspectivas interesantes en este sentido.

- Implementar una interfaz de usuario gráfica y amigable para el control del ángulo de direccionalidad del arreglo de altavoces, queda como trabajo para realizar, toda vez que actualmente esto se realiza desde la línea de comandos a través de especificar el número de muestras de retraso de la señal procesada.
- Puede asimismo extenderse el número de canales de señal de audio procesados, de modo que otros sistemas de reproducción de audio comerciales (surround 5.1, 7.1, etc.) puedan ser procesados bajo el mismo principio de independencia de la posición.
- El número de altavoces del arreglo también puede ser aumentado, de modo que la zona de audición sea mayor y puede encontrarse un número óptimo que cumpla con un compromiso adecuado de calidad y mejora respecto a la estereofonía normal, y al mismo tiempo, con la economía deseable para aplicaciones domésticas.
- El sistema de reproducción de Síntesis de Frente de Onda (WFS) es un sistema de alta calidad para la reproducción de características espaciales, de localización y carácter inmersivo. Comparado con otras técnicas, permite eliminar problemas en la percepción derivados de zonas de audición privilegiadas. El número de altavoces requeridos para implementarse hace que la técnica no sea viable para uso doméstico.
- Implementar reproducciones en WFS queda también como trabajo pendiente, sea a partir de mezclas estereofónicas comerciales o para fuentes virtuales. Trabajando con un arreglo de altavoces más numeroso que el empleado en esta tesis, se estima daría buenos resultados, según lo expuesto en el presente trabajo.

6. Bibliografia

1. McKnight, John. Why Stereo? The Philosophy of Multichannel Recording of Music. 1960. Vol 8. No. 2. Journal of the Audio Engineering Society.
2. Guastavino Catherine, Larcher Veronique, Catusseau Guillaume, Boussard Patrick. Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo. In: 13th International Conference on Auditory Display - ICAD 2007, pp. 53-59, Montréal, QC, Canada.
3. Guastavino, C. & Katz, B. Perceptual evaluation of multidimensional spatial audio reproduction. JAES. Vol. 116, No. 2, pp. 1105-1115, 2004.
4. S. Choisel & Koivuniemi. Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. JAES. Vol. 121, No. 1, pp 388-400, 2007.
5. Pedersen T. H. and Zacharov, N. How many psycho-acoustic attributes are needed? The Journal of the Acoustical Society of America, vol. 123, issue 5, p. 3163.
6. Irwan R. and Aarts Ronald. Two to Five channel sound processing. *J. Audio Eng. Soc.*, 50(11):914-926, November 2002.
7. Hull, Joseph. Dolby Laboratories Inc. Surround Past, Present and Future. A history of multichannel audio from mag strip to Dolby Digital. http://www.dolby.com/uploadedFiles/zz-Shared_Assets/English_PDFs/Professional/2_Surround_Past.Present.pdf
8. Rodenas, J. A.; Aarts R. M. Position Independent Stereophonic Sound Reproduction. *Nederlands Akoestisch Genootschap NAG Journal*, 37-47. 2001
9. Rodenas, J. A.; Aarts R. M. Sweet spot widening for stereophonic sound reproduction. IEEE WASPAA, New Paltz, NY, 21-24. October 2001, pp. 191-194.
10. Rodenas, J. A.; Aarts R. M. Derivation of an optimal directivity pattern for sweet spot widening in stereo sound reproduction. JAES Vol. 113. No. 1, January 2003.

11. Kyriakakis Chris. Fundamental and Technological Limitations of Immersive Audio Systems. Proceedings of the IEEE. Vol. 86, No. 5. May 1998.
12. Snow, William. Basic Principles of Stereophonic Sound. Society of Motion Picture and Television Engineers. Journal of SMPTE, Vol. 61, p. 567-587; Nov. 1953.
13. Berkhout, A. J; De Vries, D. Acoustic Control by Wave Field Synthesis. JAES. Vol. 93, pp. 2765-2778, 1993.
14. Berkhout, A. J. A holographic approach to acoustical control. JAES 36, 979-995, 1988.
15. Berkhout, A. J. Applied Seismic Wave Theory. Elsevier, Amsterdam 1987.
16. Wittek, H. Perception of Spatially Synthesized Sound Fields. www.hauptmikrofon.de/wfs.htm
17. Bleda Perez, Sergio; Lopez José, Escolano José. Simulación y Análisis de Prestaciones de Configuraciones de Arrays para Wave Field Synthesis. Proceedings of Symposium Nacional de la Unión Científica Internacional de Radio, La Coruña, 2003
18. Pinto, Francisco; Vetterli, Martin. Wave Field Coding in the Spacetime Frequency Domain. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, USA, 2008
19. Cobos, M; Lopez J.; Gonzalez; Escolano. Stereo to WFS Music Up-mixing: An Objective and Subjective Evaluation. *Invited Paper*, Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP 2008), St. Julians (Malta), 2008
20. Boone, Marinus. "Acoustic rendering with WFS". ACM SIGGRAPH and EUROGRAPHICS, Campfire, Snowbird, Utah. May 26-29. 2001.
21. Pueo Ortega, Basilio. Analysis and Enhancement of Multiactuator Panels for WFS Reproduction. Tesis Doctoral. Valencia Julio 2008.
22. <http://www.sea-acustica.es/Terrassa05/RDO008.pdf>
23. http://www.arsxxi.com/pfw_files/cma/ArticulosR/BoletinAELFA/2007/01/10610700060017.pdf
24. Blauert, J. Spatial Hearing: The psychophysics of Human Sound Localization. Cambridge. MA: MIT Press, 1997.

25. Blumlein, A. D. improvements in and relating to sound transmission, sound-recording and sound-reproducing systems, U. K. Patent 394 325, 1931.
26. Genelec Oy. 2004. Resolution Magazine, Sweet Spot. Pp 58-59. Vol. V3.1
27. Torick, Emil. Highlights in the History of Multichannel Sound. JAES Vol. 48, No. 1/2. 1998.
28. Hilliard, John. Acoustic Measurements on a Home Stereo Instalation. IRE Transactions on Audio. 1961.
29. Ahrens, Jens; Spors, Sascha. Implementation of Directional Sources in Wave Field Synthesis. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, USA, October 2007.
30. Ahrens, Jens; Spors, Sascha. Reproduction of a Plane Wave Sound Field using Planar and Linear Arrays of Loudspeakers. In Third IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing, March 2008.
31. Horbach; Corteel; Pellegrini; Hulsebos. Real-Time Rendering of Dynamic Scenes using Wave Field Synthesis. Proceedings of the IEEE International Conference on Multimedia, ICME 2002, Lausanne, Switzerland, August 2002.
32. Lilis, Georgios; Servetto, Sergio. dWFS: Distributed Wave Field Synthesis. Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006.
33. Gauthier, Philippe; Berry, Alain. Adaptive Wave Field Synthesis with Independent Radiation Mode Control for Active Sound Field Reproduction: Theory. JAES. Vol. 119, No. 5. May 2006.
34. Gauthier, Philippe; Berry, A; Woszyk, W. Introduction aux Fondements, aux Technologies et aux Applications Potentielles de la Synthèse de Champs Acoustiques pour la Spatialization Audio sur Reseaux de Haut-Parleurs. Harvest Moon Symposium on multichannel sound, Concordia University, 2004.
35. Toole, Floyd. The Acoustical Design of Home Theaters. Infinity, White Papers. August 1999.
36. Pulkki, Ville; Hirvonen Toni. Localization of Virtual Sources in Multichannel Audio Reproduction. IEEE transactions on Speech and Audio Processing. V. 13, January 2005.

37. De Vries, Diemer. Sound Reinforcement by Wavefield Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics. *JAES*, Vol. 44, No. 12, December 1996.
38. Romoli, Peretti, Cecchi, Palestini, Piazza. Real-Time Implementation of Wave Field Synthesis for Sound Reproduction Systems. Asia Pacific Conference on Circuits and Systems, November 30 - December 3, 2008 - Macao, China.
39. Brown, Jim. Systems for Stereophonic Sound Reinforcement: Performance Criteria, Design Techniques, and Practical Examples. AES 113th Convention, Los Angeles, CA, USA. October. 2002
40. AES staff writer. Multichannel Audio Systems and Techniques. *JAES* Vol. 53, No. 4, April, 2005.
41. C. Guastavino, Katz, B., Polack, J-D., Levitin, D., & Dubois, D. "Ecological validity of soundscape reproduction," *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 333–341, 2005.
42. Randall, R. B. Frequency Analysis. Brüel & Kjaer, 3a. Edición. 1987.
43. Bellanger, Maurice. Digital Processing of Signals. Theory and Practice. Wiley & Sons. 2nd Edition. 1988.
44. Martín Escalante Marco Vinicio. Estudio de la producción de Fuentes virtuales sonoras mediante el método de síntesis de campos sonoros. Tesis de Maestría. UNAM, México D. F. 2007.
45. Springer. Rossing, Thomas. Editor. Handbook of Acoustics. Stanford University. 2007.
46. Oppenheim, R. Schafer, and J. R. Buck. Discrete-Time Signal Processing. Prentice Hall, 2nd edition, 1999.
47. Bauer, B. B. Spatial High Fidelity. *IEEE Transactions on broadcasting*. V. BC-23. No. 3, September 1977.
48. W. E. Garrity; Hawkins J. N. A. Fantasound. Society of Motion Picture and Television Engineers. *Journal of SMPTE*, Vol. 37; Aug. 1941.
49. Dolby Laboratories Inc. The evolution of Dolby Film Sound. http://www.dolby.com/uploadedFiles/zz-Shared_Assets/English_PDFs/Professional/53_EvolutionOfSound.pdf

50. Dolby Laboratories Inc. 5.1 Channel Production Guidelines. Issue 1 S00/12957. 2000. http://www.dolby.com/uploadedFiles/zz-Shared_Assets/English_PDFs/Professional/L.mn.0002.5.1guide.pdf
51. AES Technical Council. Document AESTD1001.1.01-10. Multichannel surround Sound Systems and Operations.
52. Lifonsky, S; Colburn; Yost; Guzman. The Precedence Effect. JASA, V. 106, No. 4; October 1996.
53. Wilson, Kevin. Learning the Precedence Effect. Initial Real-World Test. IEEE Workshop on Applications of signal processing to audio and acoustics.
54. Harris; Flanagan, J; Watson. Binaural Interaction of a click pair. JASA, V. 35, 1963.
55. Bleda, S; Lopez, J; Pueo, B. Software for Simulation Performance Analysis and Real-Time implementation of WFS for 3-D Audio. Proc. Of the 6th Conference on Digital Audio Effects. London, UK. September 2003.
56. Verheijen, E. N. G. Sound Reproduction by Wave Field Synthesis. PhD Thesis. TU Delft, 1997.
57. S. Spors, H. Teutsch, and R. Rabenstein, "High-Quality Acoustic Rendering with Wave Field Synthesis," *Proc. Vision, Modeling, and Visualization Workshop*, Aka GmbH, 2002, pp. 101-108.
58. Stankovic, John; Ramamritham, Krithi. Hard Real Time Systems. IEEE Computer Society Press. 1988, Chapter 1.
59. Kuo, S.M; Lee, B. H; Thian, W. Real Time Digital Signal Processing. John Wiley & Sons, Ltd. 2006 Chapter 1.
60. Boone, M. Sound Reproduction Based in Wave Field Synthesis. AES 95th Convention, October 1993.

7. Glosario

Dipolo. Dos monopolos acústicos de igual magnitud pero fases opuestas que están colocados por una distancia pequeña d ($kd \ll 1$, siendo k el número de onda).

Filtro peine. Un filtro que se produce como consecuencia de la interferencia constructiva y destructiva de la suma de una señal con una versión retrasada de sí misma.

Frecuencia espacial de alias. Máxima frecuencia que puede ser registrada (o reproducida) en virtud de la separación espacial de los transductores de un arreglo implementado para grabación o reproducción.

HDTV. Sistema de televisión de alta definición (High Definition Television).

HRTF. Función de transferencia asociada a la cabeza. Una función de transferencia que reúne el conjunto de efectos sobre el sonido que llega a los oídos de un escucha, producidos por el torso, cabeza, y pabellón da oreja.

ITD. Diferencia Interaural de tiempo (Interaural Time Difference), diferencia de tiempo que tienen las señales que llegan a los oídos de un escucha debidas a las diferentes distancias que deben recorrer para arribar a cada uno de ellos.

ILD. Diferencia Interaural de nivel (Interaural Level Difference). Diferencia de nivel (amplitud) que tienen las señales que llegan a los oídos de un escucha.

LFE. (Efectos de Baja Frecuencia) es un canal del sistema Dolby Digital 5.1 que lleva un contenido espectral de baja frecuencia (3-120 Hz).

Monopolo. Una fuente acústica que propaga un campo de sonido que es igual en todas direcciones. Una fuente puntual.

Sistema estéreo. Sistema de reproducción de sonido más popular en el mercado durante los últimos 50 años, que basa su funcionamiento en la creación de imágenes fantasmas.

Surround. Sistemas Envoltentes. Sistemas de reproducción de sonido en los que el escucha es rodeado de altavoces y en los que se asume que las percepciones son más realistas

Sweet spot. Área de escucha óptima de los sistemas de reproducción de sonido. Utilizado, en el presente trabajo, para describir la limitada área de escucha donde el campo sonoro es reproducido correctamente.

WFS. Siglas del sistema de Síntesis de Frentes de Onda (Wave Field Synthesis). Un sistema de reproducción de sonido inmerso basado en el principio de Huygens para la construcción de campos acústicos a través de arreglos de altavoces.

8.Apéndice

8.1. Código de los programas implementados

FRENTDONDA.M

```
%PROGRAMA QUE MUESTRA GRAFICAMENTE LA PROPAGACION DE FRENTES DE ONDA,
%DEBIDOS A UNA O VARIAS FUENTES
%HUGO EZEQUIEL GUTIERREZ HERNANDEZ, 2008.

n=8; %n?mero de fuentes
d=0.5; %separaci?n entre fuentes
d1=d/2;
c=320; %velocidad de sonido
f=600;
lambda=c/f; %longitud de onda
t=2; %tiempo, posiblemente se maneje un vector
w=2*pi*(c/lambda); %2*pi*f
k=2*pi/lambda; %n?mero de onda
l=20; %del ?rea de visualizaci?n
paso=0.3; %resoluci?n de la malla de c?lculo
distx=60*lambda; %longitud de observaci?n en el eje y
disty=10*lambda; %relativo a longitud de observaci?n en x
inc=2*pi/4; %incremento de la fase
rs=zeros(1,n);
    for j=1:n
        rs(j) = -n*d/2 + d/2 + (j-1)*d;
    end
[x,y]=meshgrid(-distx:paso:distx,lambda:paso:disty);%para empezar a graficar a partir de cierto
numero de lambdas
[Nx,Ny] = size(x);
ri=zeros(Nx,Ny,n);
    for m=1:n
        ri(:, :,m)=sqrt((x-rs(m)).^2+y.^2);
    end
    %for j=1:n
    % figure
    % meshc(x,y,ri(:, :,j));
    %end
%la fase
fas=zeros(size(1,n));
    for m=1:n
        fas(m)=(m-1)*inc;
    end
    fase=exp(-i*fas);

%las fuentes
p=zeros(Nx,Ny);
    for m=1:n
        f1=(fase(m))*exp(i*(w*t-k.*ri(:, :,m)))./ri(:, :,m);
        p=p+f1;
    end
%gr?fica de resultados
close all
meshc(x,y,real(p));
colorbar
title(['FRENTE DE ONDA PRODUCIDOS POR ',num2str(n),' FUENTES SENOIDALES DE
',num2str(round(f)), ' [Hz]'], 'fontsize',12,'fontname','arial');
```

```

xlabel('POSICION DE LAS FUENTES [m]');
ylabel('DISTANCIA [m]');
figure
contourf(x,y,real(p));
colorbar
title(['FRENTE DE ONDA PRODUCIDOS POR ',num2str(n),' FUENTES SENOIDALES DE
',num2str(round(f)),' [Hz]'],'fontsize',12,'fontname','arial');
xlabel('POSICION DE LAS FUENTES [m]');
ylabel('DISTANCIA [m]');
figure
meshc(x,y,20*log10(real(p)));
colorbar
title(['FRENTE DE ONDA PRODUCIDOS POR ',num2str(n),' FUENTES SENOIDALES DE
',num2str(round(f)),' [Hz] Amplitud en [dB]'],'fontsize',12,'fontname','arial');
xlabel('POSICION DE LAS FUENTES [m]');
ylabel('DISTANCIA [m]');
figure
contourf(x,y,20*log10(real(p)));
colorbar
title(['FRENTE DE ONDA PRODUCIDOS POR ',num2str(n),' FUENTES SENOIDALES DE
',num2str(round(f)),' [Hz] Amplitud en [dB]'],'fontsize',12,'fontname','arial');
xlabel('POSICION DE LAS FUENTES [m]');
ylabel('DISTANCIA [m]');
pause
close all

```

CIRCULACION.C

```

//PROGRAMA QUE REALIZA LA DISTRIBUCION DE DIFERENTES ENTRADAS A LAS
DIFERENTES SALIDAS//
//PONDERANDOLAS Y AGREGANDO RETARDOS A ELLAS//
//2009 HUGO E.

```

```

#include "stdio.h"
#include "portaudio.h"
#define FRAMES_PER_BUFFER (256)
#define INPUTS (1)
#define OUTPUTS (4)
float entrada [INPUTS][2*FRAMES_PER_BUFFER];
float salida [OUTPUTS][2*FRAMES_PER_BUFFER];
float amplitud[OUTPUTS][INPUTS];
int retardo[OUTPUTS][INPUTS];
int N = FRAMES_PER_BUFFER;
int r, s;
int q=0;
int t=0;

/* This will be called asynchronously by the PortAudio engine. */
static int myCallback( void *inputBuffer, void *outputBuffer,
                      unsigned long framesPerBuffer, PaTimestamp outTime, void *userData )
{
    float *out = (float *) outputBuffer;
    float *in = (float *) inputBuffer;
    int n = framesPerBuffer;
    float leftInput, rightInput;
    int i, k, m, r, s;
    //int t=0;

    if( inputBuffer == NULL ) return 0;

    for (i=0; i<n; i++)
    {
        for (k=0; k<INPUTS; k++)

```

```

        {
            entrada[k][i] = entrada [k][N+i]; // se recorren las nuevas a viejas
            entrada[k][N+i] = *in++; // y se colocan en su lugar las nuevas entradas
        }
    }

    /* recorrido (amplitud, t);
    t++;
    if (t >7)
    {
        t = 0;
    }*/

    for (i=0; i<n; i++)
    {
        for (k=0; k<OUTPUTS; k++)
        {
            //salida[k][i] = salida[k][N+i]; // se colocan las salidas anteriores en su
nueva posicion
            salida[k][i] = salida[k][N+i] + 0.9999 * salida[k][N+i-1]; // se colocan las
salidas anteriores en su nueva posicion
            salida[k][N+i] = 0;
        }
    }

    for (i=0; i<n; i++)
    {
        for (k=0; k<OUTPUTS; k++)
        {
            int k_q = (k + q) % OUTPUTS; // residuo de dividir entre OUTPUTS

            /*while (k_q >= OUTPUTS)
                k_q = k_q - OUTPUTS;*/
            for (m=0; m<INPUTS; m++)
            {
                salida[k][N+i] += amplitud[k_q][m]*(entrada[m][N+i-
retardo[k_q][m]]); // multiplicacion de las entradas por sus ponderaciones

                // correspondientes
            }
        }
    }

    for (i=0; i<n; i++)
    {
        for (k=0; k<OUTPUTS; k++)
        {
            /*out++ = salida[k][N+i]; // y se envian al bufer de salida
            *out++ = (salida[k][N+i] + salida[k][N+i-1] + //; // y se envian al bufer de
salida
                salida[k][N+i-2] + salida[k][N+i-3] +
                salida[k][N+i-4] + salida[k][N+i-5] +
                salida[k][N+i-6] + salida[k][N+i-7]) / 8;
        }
    }

    t = t + 1;
    if (t >= 1024) {
        q = (q - 1) % OUTPUTS;
        t = 0;
    }
    /*while (q >= OUTPUTS)
        q = q - OUTPUTS;*/

return 0;

```

```

}
/* Open a PortAudioStream to input and output audio data. */
int main(void)
{
//matriz de amplitudes

for (r=0; r<INPUTS; r++)
{
    for (s=0; s<OUTPUTS; s++)
    {
        if (r==s)
        {
            amplitud[r][s]=1.0;
        }
        else
        {
            amplitud[r][s]=0.0;
        }
        printf(" %f",amplitud[r][s]);
    }
    printf("\n");
}

// MATRIZ DE RETARDOS
for (r=0; r<INPUTS; r++)
{
    for(s=0; s<OUTPUTS; s++)
    {
        retardo[r][s] = 0*(r*s*N)/(INPUTS*OUTPUTS);
        //printf(" %d", retardo[r][s]);
    }
    //printf("\n");
}

PortAudioStream *stream;
Pa_Initialize();
Pa_OpenDefaultStream(
    &stream,
    INPUTS, OUTPUTS,      /* entradas, salidas */
    paFloat32,
    4 * 44100.0,
    FRAMES_PER_BUFFER, 0,      /* 64 frames per buffer, let PA determine numBuffers */
    myCallback, NULL );
Pa_StartStream( stream );
getchar();
//Pa_Sleep( 100000 ); /* Sleep for 10 seconds while processing. */
Pa_StopStream( stream );
Pa_CloseStream( stream );
Pa_Terminate();
return 0;
}

```

RETIZQDER.C

```

//PROGRAMA QUE RETARDA UNA SE„AL DE ENTRADA SECUENCIALMENTE A VARIOS
ALTAVOCES (OUTPUTS)
//MEDIANTE LA MULTIPLICACION POR EXPONENCIALES COMPLEJAS (RETARDOS). EL
RETRASO SE EFECTUA HACIA
//2009 HUGO E.

#include "stdio.h"
#include "portaudio.h"
#include <fftw3.h>
#include <math.h>

```



```

#define FRAMES_PER_BUFFER (512)
#define INPUTS (1)
#define OUTPUTS (8)
int N = FRAMES_PER_BUFFER;
int NB = 2*FRAMES_PER_BUFFER;
float entrada [INPUTS][2*FRAMES_PER_BUFFER];
float salida [OUTPUTS][2*FRAMES_PER_BUFFER];
fftw_plan fft_x;
fftw_plan ifft_y;
double x[2*FRAMES_PER_BUFFER];
double y[2*FRAMES_PER_BUFFER];
double z[2*FRAMES_PER_BUFFER];
int delay = FRAMES_PER_BUFFER / (2*OUTPUTS);

void
print_vector (char * s, double * x, int m)
{
    int n;
    printf ("%s = {", s);
    for (n = 0; n < m; n++)
    {
        printf (" %lg", x[n]);
    }
    printf (" }\n");
}

void
scale_vector (double s, double * x, int m)
{
    int n;
    for (n = 0; n < m; n++)
    {
        x[n] = s * x[n];
    }
}

void
hc_mult ( double *x, double *y,int n)
{
    int k;
    double real, imag;
    x[0]=x[0]*y[0]; //componente de DC (k=0)
    if((n&1)==0)// si n es par:
    {
        x[n/2]=x[n/2]*y[n/2]; // componente de nyquist (k=n/2).
    }
    for (k=1; k<(n+1)/2; k++)
    {
        real=x[k]*y[k]-x[n-k]*y[n-k];
        imag=x[k]*y[n-k]+x[n-k]*y[k];
        x[k]=real;
        x[n-k]=imag;
    }
}

void
transformacionret (double *x, double *y, double k)
{
    double pi=4.0*atan(1.0);
    int s,t,q,i;
    q=k;
    for (i=0;i<2*N;i++)
    {
        y[i]=x[i]; //se acomoda en la variable y la informacion del vector de
muestras de entrada
    }
}

```

```

double ton=q*delay;
double to1=floor(ton);
double arg = ton*2*pi/(NB);
z[0]=1;
if ((NB&1)==0)
    {
    z[NB/2]=cos(to1*pi);
    for (s=1; s<(NB)/2; s++)
        {
        z[s] = cos(arg*s);
        z[NB-s]=sin(arg*s);
        }
    }
else
    {
    for (s=1; s<(NB+1)/2; s++)
        {
        z[s] = cos(arg*s);
        z[NB-s]=sin(arg*s);
        }
    }
hc_mult(y,z,NB); // multiplicacion por una exponencial (retardos en el dominio de
la frecuencia)
scale_vector (1.0 / NB, y, NB);
fftw_execute (ifft_y); // antitransformacion
}

/* This will be called asynchronously by the PortAudio engine. */
static int myCallback( void *inputBuffer, void *outputBuffer,
    unsigned long framesPerBuffer, PaTimestamp outTime, void *userData )
{
    float *out = (float *) outputBuffer;
    float *in = (float *) inputBuffer;
    int n = framesPerBuffer;
    int i, k, t, s;
    double pi=4.0*atan(1.0);
    int ret=N/OUTPUTS;
    int retx;

    if( inputBuffer == NULL ) return 0;
    //ACOMODO DE LAS ENTRADAS
    for (i=0; i<n; i++)
    {
        for (k=0; k<INPUTS; k++)
        {
            entrada[k][i] = entrada [k][N+i]; // se recorren las nuevas a viejas
            entrada[k][N+i] = *in++; // y se colocan en su lugar las nuevas
            entradas del bœfer de adquisicion de datos
        }
    }
    // PROCESAMIENTO DE MUESTRAS
    for (k=0; k<NB; k++)
    {
        x[k] = (double) entrada[0][k]; // se ponen las muestras en el vector 'x' para
su procesamiento a frecuencia
    }
    fftw_execute (fft_x); // transformacion de muestras a frecuencia

    for (k=0; k<OUTPUTS; k++)
    {
        transformacionret(x,y,k); // se efectua el procesamiento en frecuencia; retardo
diferente para cada salida
        for (i=0; i<n; i++)
        {
            salida[k][i] = salida[k][N+i]; // se colocan las salidas anteriores en
su nueva posicion

```

```

        salida[k][N+i] = y[N/2+i]; // se ponen las entradas anteriores a la
salida
    }
}
// ENVIO A LAS SALIDAS
for (i=0; i<n; i++)
{
    for (k=0; k<OUTPUTS; k++)
    {
        *out++ = salida[k][N+i]; // y se envian al bufer de salida
    }
}
return 0;
}

/* Open a PortAudioStream to input and output audio data. */
int main(int argc, char * argv[])
{
    PortAudioStream *stream;
    if (argc > 1)
    {
        delay = atoi(argv[1]); // Lee retardo de la l'nea de comandos
    }
    if (abs(delay * (OUTPUTS-1)) > FRAMES_PER_BUFFER/2)
    {
        fprintf(stderr, "Retardo demasiado grande.\n");
        return -1;
    }
    fft_x = fftw_plan_r2r_1d (NB, x, x,
        FFTW_R2HC, FFTW_ESTIMATE);
    ifft_y = fftw_plan_r2r_1d (NB, y, y,
        FFTW_HC2R, FFTW_ESTIMATE); // | FFTW_PRESERVE_INPUT);
    Pa_Initialize();
    Pa_OpenDefaultStream(
        &stream,
        INPUTS, OUTPUTS,          /* una entrada, ocho salidas */
        paFloat32, 44100.0,
        FRAMES_PER_BUFFER, 0,    /* 64 frames per buffer, let PA determine numBuffers */
        myCallback, NULL );
    Pa_StartStream( stream );
    getchar();
    Pa_StopStream( stream );
    Pa_CloseStream( stream );
    Pa_Terminate();
    fftw_destroy_plan (fft_x);
    fftw_destroy_plan (ifft_y);
    return 0;
}

```

ESTEREO.C

```

//Programa para una entrada y salida continua de datos, de una se'nal de dos canales, y que
coloca las fuentes hacia una orilla
//u otra de un arreglo lineal de altavoces.
//2009 HUGO E.

```

```

#include "stdio.h"
#include "portaudio.h"
#include <fftw3.h>
#include <math.h>
#define FRAMES_PER_BUFFER (8*1024)
#define INPUTS (2)
#define OUTPUTS (5)

```

```

int N = FRAMES_PER_BUFFER;

```

```

int NB = 2*FRAMES_PER_BUFFER;
float entrada [INPUTS][2*FRAMES_PER_BUFFER];
float salida [OUTPUTS][2*FRAMES_PER_BUFFER];

fftw_plan fft_xd;
fftw_plan fft_xi;
fftw_plan ifft_yd;
fftw_plan ifft_yi;

double xd[2*FRAMES_PER_BUFFER];
double yd[2*FRAMES_PER_BUFFER];
double xi[2*FRAMES_PER_BUFFER];
double yi[2*FRAMES_PER_BUFFER];
double z[2*FRAMES_PER_BUFFER];
//int delay = FRAMES_PER_BUFFER / OUTPUTS;
int delayd = FRAMES_PER_BUFFER / OUTPUTS;
int delayi = FRAMES_PER_BUFFER / OUTPUTS;

void
print_vector (char * s, double * x, int m)
{
    int n;
    printf ("%s = {", s);
    for (n = 0; n < m; n++)
        {
            printf (" %lg", x[n]);
        }
    printf (" }\n");
}

void
scale_vector (double s, double * x, int m)
{
    int n;
    for (n = 0; n < m; n++)
        {
            x[n] = s * x[n];
        }
}

void
hc_mult ( double *x, double *y,int n)
{
    int k;
    double real, imag;

    x[0]=x[0]*y[0]; //componente de DC (k=0)
    if((n&1)==0)// si n es par:
        {
            x[n/2]=x[n/2]*y[n/2]; // componente de nyquist (k=n/2).
        }
    for (k=1; k<(n+1)/2; k++)
        {
            real=x[k]*y[k]-x[n-k]*y[n-k];
            imag=x[k]*y[n-k]+x[n-k]*y[k];
            x[k]=real;
            x[n-k]=imag;
        }
}

void
transformacionretd (double *xd, double *yd, double k)
{
    double pi=4.0*atan(1.0);
    int s,t,q,i;
    q=k;

```

```

        for (i=0;i<2*N;i++)
            {
                yd[i]=xd[i]; //se acomoda en la variable y la informacion del vector de
muestras de entrada
            }
        double ton=q*delayd;
        //printf ("ret = %lg", ton);
        //printf ("\n");
        double to1=floor(ton);
        double arg = ton*2*pi/(NB);
        z[0]=1;
        if ((NB&1)==0)
            {
                z[NB/2]=cos(to1*pi);
                for (s=1; s<(NB)/2; s++)
                    {
                        z[s] = cos(arg*s);
                        z[NB-s]=sin(-arg*s);
                    }
            }
        else
            {
                for (s=1; s<(NB+1)/2; s++)
                    {
                        z[s] = cos(arg*s);
                        z[NB-s]=sin(-arg*s);
                    }
            }
        hc_mult(yd,z,NB); // multiplicacion por una exponencial (retardos en el dominio
de la frecuencia)
        scale_vector (1.0 / NB, yd, NB);
        fftw_execute (ifft_yd); // antitransformacion
    }

void
transformacionreti (double *x, double *y, double k)
{
    double pi=4.0*atan(1.0);
    int s,t,q,i;
    q=OUTPUTS-1-k;
        for (i=0;i<2*N;i++)
            {
                yi[i]=xi[i]; //se acomoda en la variable y la informacion del vector de
muestras de entrada
            }
        double ton=q*delayi;
        //printf ("ret = %lg", ton);
        //printf ("\n");
        double to1=floor(ton);
        double arg = ton*2*pi/(NB);
        z[0]=1;
        if ((NB&1)==0)
            {
                z[NB/2]=cos(to1*pi);
                for (s=1; s<(NB)/2; s++)
                    {
                        z[s] = cos(arg*s);
                        z[NB-s]=sin(-arg*s);
                    }
            }
        else
            {
                for (s=1; s<(NB+1)/2; s++)
                    {
                        z[s] = cos(arg*s);
                        z[NB-s]=sin(-arg*s);
                    }
            }
}

```

```

    }
    hc_mult(yi,z,NB); // multiplicacion por una exponencial (retardos en el dominio de
la frecuencia)
    scale_vector (1.0 / NB, yi, NB);
    fftw_execute (ifft_yi); // antitransformacion
}

/* This will be called asynchronously by the PortAudio engine. */
static int myCallback( void *inputBuffer, void *outputBuffer,
    unsigned long framesPerBuffer, PaTimestamp outTime, void *userData )
{
    float *out = (float *) outputBuffer;
    float *in = (float *) inputBuffer;
    int n = framesPerBuffer;
    int i, k, t, s;
    double pi=4.0*atan(1.0);

    if( inputBuffer == NULL ) return 0;
    //ACOMODO DE LAS ENTRADAS
    for (i=0; i<n; i++)
    {
        for (k=0; k<INPUTS; k++) {
            entrada[k][i] = entrada [k][N+i]; // se recorren las nuevas a viejas
            entrada[k][N+i] = *in++; // y se colocan en su lugar las nuevas
entradas del búfer de adquisicion de datos
        }
        // PROCESAMIENTO DE MUESTRAS
        for (k=0; k<NB; k++)
        {
            xd[k] = (double) entrada[0][k]; // se ponen las muestras en el vector 'x'
para su procesamiento a frecuencia
            xi[k] = (double) entrada[1][k]; // lo mismo, pero para el segundo canal
        }
        fftw_execute (fft_xd); // transformacion de muestras a frecuencia, canal 1
        fftw_execute (fft_xi); // transformacion de muestras a frecuencia, canal 2

        for (k=0; k<OUTPUTS; k++)
        {
            transformacionretd(xd,yd,k); // se efectua el procesamiento en frecuencia;
retardo diferente para cada salida
            transformacionreti(xi,yi,k); // se efectua el procesamiento en frecuencia; retardo
diferente para cada salida
            for (i=0; i<n; i++)
            {
                salida[k][i] = salida[k][N+i]; // se colocan las salidas anteriores en
su nueva posicion
                salida[k][N+i] = (yd[N+i] + yi[N+i]); // se ponen las entradas
anteriores a la salida
            }
        }
        // ENVIO A LAS SALIDAS
        for (i=0; i<n; i++)
        {
            for (k=0; k<OUTPUTS; k++)
            {
                *out++ = salida[k][N+i]; // y se envian al bufer de salida
            }
        }

        return 0;
    }
}

/* Open a PortAudioStream to input and output audio data. */
int main(int argc, char * argv[])
{
    PortAudioStream *stream;

```

```

    if (argc > 1)
        {
            delayd = atoi(argv[1]);
            delayi = atoi(argv[2]); // Lee retardo de la línea de comandos
        }
    if (abs(delayd * (OUTPUTS-1)) > FRAMES_PER_BUFFER)
        {
            fprintf (stderr, "Retardo demasiado grande.\n");
            return -1;
        }
    fft_xd = fftw_plan_r2r_1d (NB, xd, xd,
                              FFTW_R2HC, FFTW_ESTIMATE);
    fft_xi = fftw_plan_r2r_1d (NB, xi, xi,
                              FFTW_R2HC, FFTW_ESTIMATE);
    ifft_yd = fftw_plan_r2r_1d (NB, yd, yd,
                               FFTW_HC2R, FFTW_ESTIMATE); // | FFTW_PRESERVE_INPUT);
    ifft_yi = fftw_plan_r2r_1d (NB, yi, yi,
                               FFTW_HC2R, FFTW_ESTIMATE); // | FFTW_PRESERVE_INPUT);

    Pa_Initialize();
    Pa_OpenDefaultStream(
        &stream,
        INPUTS, OUTPUTS,          /* una entrada, ocho salidas */
        paFloat32, 44100.0,
        FRAMES_PER_BUFFER, 0,    /* 64 frames per buffer, let PA determine numBuffers */
        myCallback, NULL );
    Pa_StartStream( stream );
    getchar();
    Pa_StopStream( stream );
    Pa_CloseStream( stream );
    Pa_Terminate();

    fftw_destroy_plan (fft_xd);
    fftw_destroy_plan (ifft_yd);
    fftw_destroy_plan (fft_xi);
    fftw_destroy_plan (ifft_yi);
    return 0;
}

```

ESTEREOM.C

// Programa que realiza el movimiento de dos fuentes de un arreglo de altavoces mediante la aplicacion de retrasos en el dominio // de la frecuencia. Ambas fuentes pueden ser movidas del centro a cada lado del arreglo. // 2009 HUGO E.

```

#include "stdio.h"
#include "portaudio.h"
#include <fftw3.h>
#include <math.h>
#define FRAMES_PER_BUFFER (4*512)
#define INPUTS (2)
#define OUTPUTS (5)

int N = FRAMES_PER_BUFFER;
int NB = 2*FRAMES_PER_BUFFER;
float entrada [INPUTS][2*FRAMES_PER_BUFFER];
float salida [OUTPUTS][2*FRAMES_PER_BUFFER];
fftw_plan fft_xi;
fftw_plan ifft_yi;
fftw_plan fft_xd;
fftw_plan ifft_yd;
double xi[2*FRAMES_PER_BUFFER];
double yi[2*FRAMES_PER_BUFFER];
double xd[2*FRAMES_PER_BUFFER];

```

```

double yd[2*FRAMES_PER_BUFFER];
double z[2*FRAMES_PER_BUFFER];
int delayi = FRAMES_PER_BUFFER / (2*OUTPUTS);
int delayd = FRAMES_PER_BUFFER / (2*OUTPUTS);

void
print_vector (char * s, double * x, int m)
{
    int n;
    printf ("%s = {", s);
    for (n = 0; n < m; n++)
    {
        printf (" %lg", x[n]);
    }
    printf (" }\n");
}

void
scale_vector (double s, double * x, int m)
{
    int n;
    for (n = 0; n < m; n++)
    {
        x[n] = s * x[n];
    }
}

void
hc_mult ( double *x, double *y,int n)
{
    int k;
    double real, imag;
    x[0]=x[0]*y[0]; //componente de DC (k=0)
    if((n&1)==0)// si n es par:
    {
        x[n/2]=x[n/2]*y[n/2]; // componente de nyquist (k=n/2).
    }
    for (k=1; k<(n+1)/2; k++)
    {
        real=x[k]*y[k]-x[n-k]*y[n-k];
        imag=x[k]*y[n-k]+x[n-k]*y[k];
        x[k]=real;
        x[n-k]=imag;
    }
}

void
transformacionreti (double *x, double *y, double k)
{
    double pi=4.0*atan(1.0);
    int s,t,q,i;
    q=k;
        for (i=0;i<2*N;i++)
        {
            yi[i]=xi[i]; //se acomoda en la variable y la informacion del vector de
muestras de entrada
        }
        double ton=q*delayi;
        //printf ("ret = %lg", ton);
        //printf ("\n");
        double to1=floor(ton);
        double arg = ton*2*pi/(NB);
        z[0]=1;
        if ((NB&1)==0)
        {
            z[NB/2]=cos(to1*pi);

```



```

        for (s=1; s<(NB)/2; s++)
            {
                z[s] = cos(arg*s);
                z[NB-s]=sin(arg*s);
            }
        }
        else
        {
            for (s=1; s<(NB+1)/2; s++)
                {
                    z[s] = cos(arg*s);
                    z[NB-s]=sin(arg*s);
                }
        }
        hc_mult(yi,z,NB); // multiplicacion por una exponencial (retardos en el dominio de
la frecuencia)
        scale_vector (1.0 / NB, y, NB);
        fftw_execute (ifft_yi); // antitransformacion
    }

void
transformacionretd
(double *x, double *y, double k)
{
    double pi=4.0*atan(1.0);
    int s,t,q,i;
    q=k;
        for (i=0;i<2*N;i++)
            {
                yd[i]=xd[i]; //se acomoda en la variable y la informacion del vector de
muestras de entrada
            }
        double ton=q*delayd;
        //printf ("ret = %lg", ton);
        //printf ("\n");
        double to1=floor(ton);
        double arg = ton*2*pi/(NB);
        z[0]=1;
        if ((NB&1)==0)
            {
                z[NB/2]=cos(to1*pi);
                for (s=1; s<(NB)/2; s++)
                    {
                        z[s] = cos(arg*s);
                        z[NB-s]=sin(arg*s);
                    }
            }
        else
        {
            for (s=1; s<(NB+1)/2; s++)
                {
                    z[s] = cos(arg*s);
                    z[NB-s]=sin(arg*s);
                }
        }
        hc_mult(yd,z,NB); // multiplicacion por una exponencial (retardos en el dominio
de la frecuencia)
        scale_vector (1.0 / NB, y, NB);
        fftw_execute (ifft_yd); // antitransformacion
    }

/* This will be called asynchronously by the PortAudio engine. */
static int myCallback( void *inputBuffer, void *outputBuffer,
    unsigned long framesPerBuffer, PaTimestamp outTime, void *userData )
{
    float *out = (float *) outputBuffer;

```

```

float *in = (float *) inputBuffer;
int n = framesPerBuffer;
int i, k, t, s;
double pi=4.0*atan(1.0);

if( inputBuffer == NULL ) return 0;
//ACOMODO DE LAS ENTRADAS
for (i=0; i<n; i++)
{
    for (k=0; k<INPUTS; k++)
    {
        entrada[k][i] = entrada [k][N+i]; // se recorren las nuevas a viejas
        entrada[k][N+i] = *in++; // y se colocan en su lugar las nuevas
entradas del búfer de adquisicion de datos
    }
}
// PROCESAMIENTO DE MUESTRAS
for (k=0; k<NB; k++)
{
    xi[k] = (double) entrada[0][k]; // se ponen las muestras en el vector 'x' para
su procesamiento a frecuencia
    xd[k] = (double) entrada[1][k];
}
fftw_execute (fft_xi); // transformacion de muestras a frecuencia
fftw_execute (fft_xd);

for (k=0; k<OUTPUTS; k++)
{
    transformacionreti(xi,yi,k);
    transformacionretd(xd,yd,k); // se efectua el procesamiento en frecuencia;
retardo diferente para cada salida
for (i=0; i<n; i++)
{
    salida[k][i] = salida[k][N+i]; // se colocan las salidas anteriores en
su nueva posicion
    salida[k][N+i] = yi[N/2+i] + yd[N/2+i] ; // se ponen las entradas
anteriores a la salida
}
}
// ENVIO A LAS SALIDAS
for (i=0; i<n; i++)
{
    for (k=0; k<OUTPUTS; k++)
    {
        *out++ = salida[k][N+i]; // y se envian al bufer de salida
    }
}

return 0;
}
/* Open a PortAudioStream to input and output audio data. */
int main(int argc, char * argv[])
{
    PortAudioStream *stream;
    if (argc > 1)
    {
        delayi = atoi(argv[1]);
        delayd = atoi(argv[2]); // Lee retardo de la línea de comandos
    }
    if (abs(delayi * (OUTPUTS-1)) > FRAMES_PER_BUFFER/2)
    {
        fprintf (stderr, "Retardo demasiado grande.\n");
        return -1;
    }
    fft_xi = fftw_plan_r2r_1d (NB, xi, xi,
        FFTW_R2HC, FFTW_ESTIMATE);
    fft_xd = fftw_plan_r2r_1d (NB, xd, xd,

```

```

        FFTW_R2HC, FFTW_ESTIMATE);
    ifft_yi = fftw_plan_r2r_1d (NB, yi, yi,
        FFTW_HC2R, FFTW_ESTIMATE);
    ifft_yd = fftw_plan_r2r_1d (NB, yd, yd,
        FFTW_HC2R, FFTW_ESTIMATE); // FFTW_PRESERVE_INPUT);

Pa_Initialize();
Pa_OpenDefaultStream(
    &stream,
    INPUTS, OUTPUTS,          /* una entrada, ocho salidas */
    paFloat32, 44100.0,
    FRAMES_PER_BUFFER, 0,    /* 64 frames per buffer, let PA determine numBuffers */
    myCallback, NULL );
Pa_StartStream( stream );
getchar();
Pa_StopStream( stream );
Pa_CloseStream( stream );
Pa_Terminate();

    fftw_destroy_plan (fft_xi);
    fftw_destroy_plan (ifft_yi);
    fftw_destroy_plan (fft_xd);
    fftw_destroy_plan (ifft_yd);
return 0;
}

```