



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**“DESARROLLO DE UNA TÉCNICA PARA GENERAR
AUDIO ENVOLVENTE SIMULADO”**

T E S I S

**QUE PARA OBTENER EL TÍTULO DE:
INGENIERO EN TELECOMUNICACIONES**

PRESENTA:

DANIEL LÓPEZ GARCÍA

ASESOR: DR. MIGUEL MOCTEZUMA FLORES

MÉXICO, D.F. AGOSTO 2013

AGRADECIMIENTOS

A mis padres



TABLA DE CONTENIDO

Introducción	6
1.1 Motivación	7
1.2 El oído en la historia evolutiva del hombre	7
1.3 Realidad aumentada y el sentido del oído en la actualidad	10
1.4 Ejemplos de realidad aumentada.....	13
1.4.1 Google Glass.....	14
1.4.2 Touch screens.....	15
1.4.3 E-ink.....	16
1.4.4 nVidia Tessellation	17
1.4.5 Nokia rich audio recording	18
1.4.6 MP3.....	18
1.4.7 Audio 5.1	19
1.5 Definición del problema a tratar	21
La generación de sonido envolvente en la actualidad	24
2.1 Grabación múltiple.....	25
2.1.1 Microfonía binaural.....	25
2.1.2 Grabación múltiple con reproducción múltiple	27
2.2 Principio de Huygens.....	29
2.3 Head Related Transfer Function	29
El oído humano	32
3.1 Respuesta en frecuencia	33
3.2 Rango dinámico	36
3.3 Patrón de captación	36
3.4 Fase y tiempo de llegada	42
3.5 Parámetros psicoacústicos relevantes	43
3.5.1 Latencia.....	43
3.5.2 Efecto Franssen	45
Propagación del sonido	46
4.1 El sonido como una onda.....	47
4.2 Frecuencia	48
4.3 Velocidad de propagación.....	48
4.4 Dirección de propagación	49
4.5 Longitud de onda	49
4.6 Frente de onda	50
4.7 Refracción	51
4.8 Difracción	53



INTRODUCCIÓN



4.9 Patrón de radiación.....	54
4.10 Dispersión acústica	56
El sistema oído-aire-fuente de sonido	57
5.1 Consideraciones	58
5.2 Diferencia de tiempo interaural.....	58
5.3 Fase de llegada, dirección de propagación y retraso de tiempo adicional...	62
5.4 Función de transferencia.....	64
5.5 Modelado del sistema.....	64
Implementación del modelo en un simulador de tiempo real.....	67
6.1 Elección de las tecnologías disponibles	68
6.1.1 Digital o analógico	68
6.1.2 Formatos de audio	69
6.1.3 Lenguajes de programación	70
6.1.4 Librerías para manipulación de audio.....	70
6.1.5 Descripción del sistema.....	71
6.2 Diagrama de bloques del sistema.....	74
6.3 Implementación final.....	77
Conclusiones	97
7.1 Comentarios de la implementación	98
7.2 Restricciones de la implementación.....	100
7.3 Ventajas y desventajas	100
Referencias bibliográficas	102

CAPÍTULO 1

INTRODUCCIÓN

*¿Desde cuándo el diseño de software ha sido
sobre lo que la gente quiere?
Es una simple cuestión de evolución.*

- Bill Gates



INTRODUCCIÓN



1.1 Motivación

El presente trabajo tiene por objeto dar un tratado sobre la situación actual de las tecnologías de generación de audio tridimensional, revisando de manera general como funcionan. Además, una vez llegado al punto donde se conozcan las distintas metodologías, se incluye el desarrollo de una técnica propia.

Para ello este capítulo está dedicado a hacer un repaso sobre el sentido del oído desde un punto de vista evolutivo, así como la situación actual de este sentido en su relación con el desarrollo tecnológico.

1.2 El oído en la historia evolutiva del hombre

La teoría evolutiva sostiene que aquellas especies que mejor se adaptan para sobrevivir en su ambiente son las que continuarán existiendo por encima de las que no se adapten tan bien. Esto significa que a lo largo de las generaciones, cada especie sobreviviente se perfecciona a sí misma, eliminando de su propia especie, o acervo genético, aquellas mutaciones que repetidamente fracasan en la meta última que como organismo tiene cada ser vivo que habita este planeta: reproducirse.

En el ser humano, una de las especies más exitosas del planeta, la característica que sin duda lo ha colocado en la cima de la cadena alimenticia es su capacidad intelectual, que incluye el razonamiento lógico, el uso del lenguaje, entre otras características derivadas de su capacidad intelectual.



CAPÍTULO 1



Sin embargo, en el largo camino que siguió esta especie para llegar a lo que es hoy, hubo muchas dificultades que el medio ambiente le impuso. Una de ellas, tan importante como las demás, fue el punto en el que de acuerdo a las teorías evolutivas actuales, abandonó los mares, lagos, ríos, para convertirse en un organismo terrestre.

Entre las muchas adaptaciones evolutivas que en ese punto de su evolución se dieron, se encuentra el desarrollo del oído, tal y como lo conocemos en la mayoría de los vertebrados de hoy en día, incluidas aves, reptiles y mamíferos. Este desarrollo fue muy importante durante los primeros pasos de esas antiguas especies, en particular los mamíferos pequeños, sujetos constantes de caza por parte de organismos más grandes.

Es evidente que las especies que no forman parte de la cima en la cadena alimenticia deben desarrollar todas las defensas que sean necesarias para sobrevivir, y para ello basta echar un vistazo a toda la variedad de mecanismos de defensa que se han ido desarrollando y perfeccionado a lo largo de la evolución, tal y como el camuflaje, la piel venenosa, la piel con espinas, la velocidad de huida, etc. Pero, ¿cómo saber en qué momento es prudente estar alerta y cuándo no? La percepción del entorno es quizá la característica más importante que los seres vivos pueden tener.

Entonces, ¿qué deberíamos percibir del medio ambiente? Cabe mencionar que perteneciendo a un mundo de tres dimensiones espaciales, nuestra mejor opción



INTRODUCCIÓN



es buscar percibirlo tal y como es: como un mundo tridimensional. Es por ello que cuando hablamos del sentido de la vista y del oído, la mayoría de los seres vivos que tienen estos sentidos tienen dos órganos idénticos, con una pequeña pero importante diferencia: su ubicación en el cuerpo anfitrión. Esta diferencia les permite trabajar como órganos co-referenciales para generar un mapa mental de nuestro entorno tan preciso como sea necesario. Así pues, contestando la pregunta inicial de este párrafo, cada ser vivo deberá percibir lo que sea estrictamente necesario para maximizar sus posibilidades de sobrevivencia. Significa esto que la percepción del medio ambiente que cada especie tiene no es una versión fiel de lo que realmente sucede, sino simplemente de aquellas cosas que suponen una amenaza para su existencia.

En el caso del oído, una de sus varias tareas es alertar sobre la posición relativa de una posible amenaza (o una presa, aunque muy pocos organismos dependen del oído como cazadores. Sólo seres con oídos superiores, como los murciélagos, pueden hacerlo). Así, saber en qué dirección y a qué distancia se encuentra una amenaza es la guía inicial para saber cómo actuar. Si está muy cerca, la acción es utilizar toda la energía disponible para huir tan rápido como sea posible. Si está a una mayor distancia, lo más prudente es alejarse paulatinamente, sin caer en la necesidad de desperdiciar energía para una huida rápida, reservándola para momentos en los que sea realmente necesario.

A diferencia de la vista, cuya función principal es la búsqueda, e identificación de los elementos de nuestro entorno, el oído inicialmente actuó como un radar permanente, inagotable y lo mejor posible diseñado para alertar. Por mejor posible



CAPÍTULO 1



debemos entender lo estrictamente necesario: ¿Representa una amenaza algo que se encuentra a 100 metros de distancia? Si es así, entonces deberíamos de ser más veloces, y no preocuparnos en tener un oído tan sensible. ¿Representa una amenaza un objeto por encima de nosotros? Rara vez un ser humano es presa de las aves, y aunque pueda serlo, la evolución solo dota de defensas contra las amenazas más frecuentes y con mayor peligro. Por estas razones, el oído tiene características distintas en cada especie, según sus necesidades, y veremos cuáles son estas características en el ser humano en un capítulo posterior.

1.3 Realidad aumentada y el sentido del oído en la actualidad

Pero volvamos al presente. ¿Qué sucede con el ser humano? Desde la revolución industrial (y para algunos individuos, los de la clase élite, desde que se crearon las complejas jerarquías sociales que incluían la esclavitud), muchas características que antes eran esenciales para la meta última de cada individuo se han vuelto innecesarias desde entonces. ¿Qué tan importante es la fuerza física ahora que la maquinaria hace el mayor trabajo a beneficio de la humanidad? ¿Qué tan alerta debe estar una persona que habita en una ciudad de ser asechada por un depredador salvaje, como bien pudiera haber parecido posible hace apenas algunos cientos de años?

El ser humano está cambiando las reglas del juego en cuanto a la evolución, ya que ahora está adaptando el mundo a sus necesidades, en lugar de continuar adaptándose. Es un lujo que sólo las especies dominantes pueden darse:



INTRODUCCIÓN



experimentar con su acervo genético. Y la humanidad está experimentando a modificar el mundo a lo que considera favorable para sí mismo. Medios de transporte, medios de comunicación, industria de alimentos, economías basadas en conocimiento, son apenas algunos cambios que la humanidad ha hecho a su vida, a través de la tecnología, supuestamente para facilitarse la existencia. De esta manera, sujetos que en un mundo salvaje difícilmente podrían haber sobrevivido ahora están en posibilidades de llevar una vida considerada normal e incluso hacer aportaciones valiosas a la humanidad: desde Ludwig van Beethoven, de capacidad auditiva reducida, hasta Stephen Hawking, de capacidad motriz prácticamente nula para valerse por sí mismo, estamos generando una nueva especie con un acervo genético sin precedentes, ya que las mutaciones antes consideradas inservibles ahora están probando ser útiles en este nuevo orden genético.

De entre todos estos cambios, el contacto a distancia sin duda potencializará a las sociedades del futuro, ya que el tiempo y el costo para transportarse están aumentando toda vez que las ciudades están aumentando su población, toda vez que la presencia física para comunicarnos se está tornando innecesaria, gracias a los avances en las telecomunicaciones. Es ya hoy en día una realidad que muchos trabajos se realizan desde casa. Que un porcentaje cada vez mayor de comunicación entre nuestros familiares se realiza de manera remota, ya sea mediante texto o mediante voz. Hay quienes dicen incluso que las emociones humanas, consideradas muy debajo en la estructura de nuestro sistema nervioso y por tanto muy resistentes al cambio, están viéndose modificadas por nuevas tecnologías: ¿Qué significa, por ejemplo, extrañar hoy en día? Sin duda, no es la



CAPÍTULO 1



misma emoción que se tenía hace algunos siglos, donde no se sabía absolutamente nada de las personas alejadas. En el presente, donde la distancia ya no impone un límite en la cantidad de información que podamos compartir con nuestros seres queridos, el término extrañar está comenzando a limitarse a la simple presencia física. Se está volviendo una emoción incompleta.

Con la globalización, el hombre del presente necesita más que nunca transportarse a prácticamente cualquier parte del mundo. Pero, ¿no sería más fácil simplemente interactuar a distancia con las partes envueltas?

También, con las nuevas posibilidades de exploración submarina, o de ambientes no aptos para las capacidades físicas del hombre, se están generando equipos mecánicos controlados remotamente que hagan las veces de un explorador. El problema es que siempre tienen que ser operados por un ser humano, pues carecen de inteligencia propia. ¿No sería una buena idea que el operador humano tuviera una idea más completa del ambiente que rodea a la máquina y no del ambiente que lo rodea? Esto lo transportaría virtualmente al lugar al que no puede ir por las altas presiones, bajas o altas temperaturas, ausencia de oxígeno, etc.

Por eso y por más razones, el hombre de hoy en día está trabajando para crear ambientes virtuales de comunicación y, en su afán por volver esto una forma de vida moderna, está creando lo que se conoce como realidad aumentada, que es la combinación de entornos virtuales con entornos reales, maximizando el aprovechamiento de los sentidos actuales del hombre. Todas las tecnologías de comunicación hasta finales del siglo XX eran claramente distinguibles de la



INTRODUCCIÓN



realidad, condenándolas a ser vistas como algo ajeno a nuestra naturaleza, pero hoy en día la integración que se está teniendo es cada vez mayor.

1.4 Ejemplos de realidad aumentada

Como ya se dijo en el apartado anterior, las últimas tecnologías buscan fundirse con la experiencia cotidiana de los seres humanos. A continuación se muestran algunos ejemplos rápidos para darnos una idea de adónde se está buscando llegar.



1.4.1 Google Glass

Unas pequeñas gafas con conexión a Internet capaces de interponer imágenes como si estuviéramos viendo una pantalla suspendida al frente de nosotros. Incluye servicios como acceso a redes sociales, mapas, enciclopedias en líneas, llamadas de audio y vídeo, grabación, etc.



Figura 1.1 Google glass proyecta una imagen virtual sobre lo que estamos viendo. En la imagen superior derecha se indica información de la estación de transporte público a tomar. En la imagen central derecha se proyecta un mapa sobre la calle que estamos recorriendo y en la imagen inferior derecha información de la tienda que estamos viendo.



INTRODUCCIÓN



1.4.2 Touch screens

Las pantallas táctiles han visto una evolución impresionante en los últimos años. Desde las pantallas basadas en tecnologías resistivas con su mayor auge a principios de la década de los 90s hasta las avanzadas pantallas capacitivas de principios de 2000 capaces de procesar patrones avanzados de movimientos, este elemento ha enriquecido la interacción humano-máquina al grado de haber hecho obsoletos dispositivos anteriores tales como el mouse, el teclado físico, los lápices digitales, etc. Esta tecnología, combinada con la retroalimentación háptica, le da al ser humano uno de las experiencias más reales al dar órdenes a sus equipos de cómputo móviles.



Figura 1.2 Las computadoras que cuentan con entrada táctil en la pantalla mejoran notablemente la experiencia del usuario al comportarse de una manera más natural y acorde con lo que una persona podría esperar en su mundo natural.



1.4.3 E-ink

La tinta electrónica es otra prueba del paso que está dando el ser humano para integrar lo artificial con lo natural. Enfrentado con las limitaciones actuales de las pantallas existentes, basadas en tubos de rayos catódicos, cristal líquido, plasma, LED o proyecciones, pues todas estas formas de presentación visual tienen en común el hecho de ser fuentes de luz para poder presentar al ojo humano información visual, cosa que en la naturaleza rara vez sucede. La tinta electrónica, por el contrario, acomoda las moléculas en una delgada capa líquida, cada molécula con un color diferente, para reflejar, no emitir, la luz del medio ambiente. Esta importante diferencia le da la cualidad de percibirse como un elemento impreso, o como cualquier otro objeto natural de nuestro medio ambiente: Uno que refleja la luz.

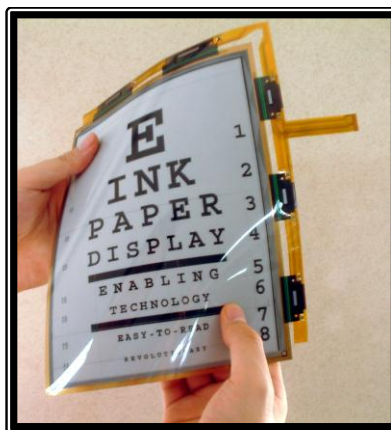


Figura 1.3 Pantalla de tinta electrónica flexible. Capaz de modificar la información que muestra, tiene la apariencia de ser una hoja de papel impresa con tinta tradicional.



1.4.4 nVidia Tessellation

Al día de hoy, las unidades de procesamiento gráfico que pueden adquirirse en la gama alta de productos de nVidia son capaces de producir, en tiempo real, rostros y objetos de nuestra vida cotidiana, simulando la física detrás de ellos. En el caso de los rostros humanos, son capaces de simular la contracción y distensión de los músculos de la cara, dando un realismo nunca antes visto:



Figura 1.4 Rostro humano completamente generado por una computadora de uso comercial de entretenimiento con un chip gráfico nVidia. El chip gráfico calcula los reflejos visibles en todo el rostro, la translucidez de la oreja en el lóbulo superior, el brillo de los ojos, con un nivel de detalle que es indistinguible de un rostro real. El movimiento del rostro, asimismo, es muy natural.



CAPÍTULO 1



En el campo del audio, no podemos dejar de mencionar algunos ejemplos:

1.4.5 Nokia rich audio recording

Tecnología de grabación de audio integrada en los últimos celulares de la compañía finlandesa Nokia, permite grabar el mismo rango dinámico que el oído humano (130 dB), en un rango de frecuencia también igual al oído humano (20 Hz – 20kHz). Además, compuesto de dos micrófonos opuestos en los dispositivos móviles que integran esta tecnología, genera audio binaural que enriquece la experiencia al reproducir el ambiente grabado de una manera más fiel, de tal manera que al reproducirlo puede repetirse la sensación de “estar físicamente presente”.

1.4.6 MP3

Uno de los mayores avances en el campo del audio, es uno de los primeros métodos de compresión de audio que no parte de las bases físicas del oído, ni de las conocidas metodologías en teoría de la información como herramienta principal para reducir el tamaño de un archivo de audio. En lugar de eso, parte del funcionamiento neuronal del sistema de percepción del sonido, es decir, del modelo psicoacústico, para así dar pie a la eliminación de información auditiva innecesaria, para fines prácticos, que permitan reducir el tamaño final de un archivo de audio hasta 10 veces sin pérdida perceptible en la calidad. Destaca por



INTRODUCCIÓN



ser una tecnología que llevó a cabo un entendimiento profundo de la percepción auditiva para así producir avances importantes en la industria del audio.

1.4.7 Audio 5.1

Considerado como el punto culminante de la reproducción de audio, la reproducción de 6 canales provee la mejor experiencia auditiva actual en cuanto a audio se refiere, entregándole a los espectadores la mayor cantidad de información sobre la posición de las fuentes virtuales de audio como sea posible.

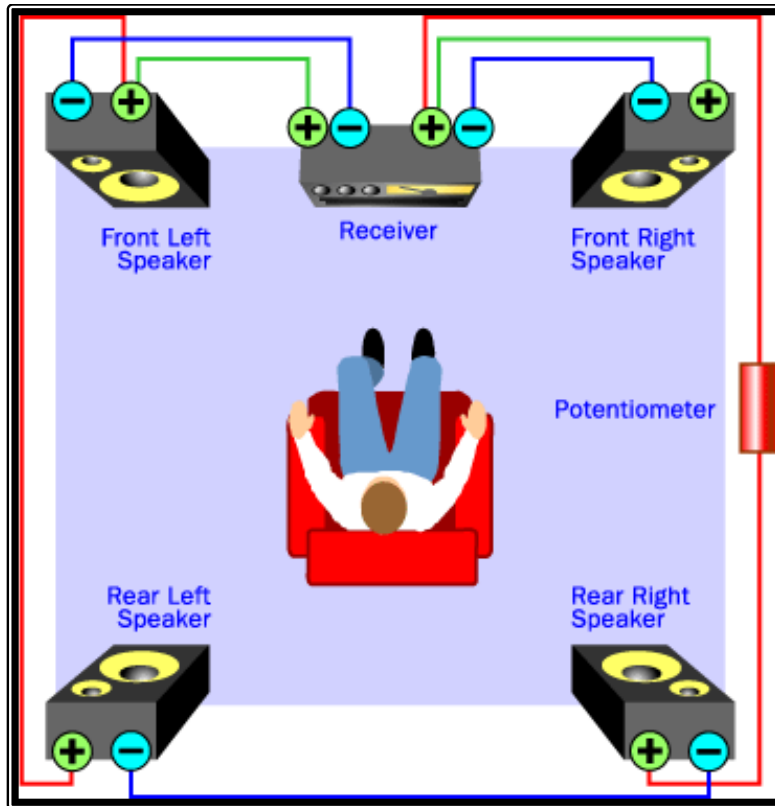


Figura 1.5 Simulación de audio envolvente a partir de la colocación física de varias fuentes de sonido, que al combinarse son capaces de producir varios efectos tridimensionales en la experiencia auditiva del espectador.

Si bien ha habido una cantidad importante de avances en el campo de las tecnologías de audio, el oído, al ser considerado un sentido ligeramente menos importante que la vista al menos para el ser humano, ha visto disminuida la cantidad de avances que en su campo se hacen a favor de las tecnologías visuales. Prueba práctica de ello es la existencia de procesadores gráficos normalmente más potentes que los procesadores centrales en una computadora



INTRODUCCIÓN



contra procesadores de audio menos poderosos e incluso con frecuencia implementados en software.

En este sentido, un factor que no se ha resuelto es el de nuestra percepción auditiva tridimensional, la cual no se ha emulado o dado una solución aceptable que sea compatible con el uso de auriculares. Es este factor la motivación del presente trabajo.

1.5 Definición del problema a tratar

Una de las mayores carencias que se encuentra en el mundo tecnológico actual es la capacidad de reproducir sonido que para fines de la percepción humana, formen parte del ambiente en donde se encuentra, es decir, sonido envolvente generado en tiempo real.

Esta carencia impone restricciones en las tecnologías actuales ya que al no poder definir la posición de una fuente virtual de sonido la experiencia buscada está condenada a ser percibida como artificial, fracasando en el objetivo actual de fundir las tecnologías que interactúan con el ser humano con el ambiente que lo rodea.

Es por eso que el presente trabajo pretende explorar las tecnologías actuales para lograrlo, en el capítulo 2, y además aportar una técnica que permita dar una solución a ese problema, en el capítulo 6, de tal manera que al integrarse en los dispositivos actuales fomente la percepción de una realidad aumentada,



CAPÍTULO 1



enriquecida con entornos virtuales pensados para mejorar la forma de vida de las personas.

Pero para llegar a esa técnica propia, primero se hará un estudio de cómo percibimos los sonidos de nuestro entorno para así poder explorar las diferentes posibilidades de simularlo con las herramientas actuales. Para este fin, el capítulo 3 estará dedicado a un estudio de las características del oído humano, el capítulo 4 estudiará la propagación del sonido con un enfoque sobre su relación con el oído. El capítulo 5 unirá las ideas adquiridas de los capítulos 3 y 4, para combinar todos los fenómenos que intervienen en la percepción tridimensional del sonido, para así modelar un sistema desde un punto de vista sintetizado, es decir, obtendremos un sistema descrito funcionalmente, el cuál conteste a la pregunta ¿qué se tiene que hacer a una señal de sonido para que se perciba como si tuviera información sobre su posición relativa al escucha? El capítulo 6, explorará la manera de lograrlo, es decir, contestará la pregunta ¿cómo lograrlo?.

Finalmente, en el capítulo 7, se comentará la técnica dada comparándola con las actuales, así como vislumbrando sus fortalezas y debilidades.



INTRODUCCIÓN



Antes de continuar, es necesario definir claramente en este punto lo que se pretende hacer de manera general:

Para aplicaciones o situaciones en las que se tenga una fuente de sonido grabada en formato monaural, será posible reproducirla en auriculares (binaurales) con modificaciones que permitan brindarle información al cerebro de la ubicación virtual del sonido, entendiendo por tiempo real como la capacidad de modificar dicha posición percibida de manera arbitraria obteniendo una respuesta instantánea sobre el cambio de la posición. Para esta situación definiremos lo que se entiende por instantáneo en un capítulo posterior.

CAPÍTULO 2

LA GENERACIÓN DE SONIDO ENVOLVENTE EN LA ACTUALIDAD

*Se ha vuelto obvio que
nuestra tecnología supera nuestra humanidad*
- Albert Einstein



LA GENERACIÓN DE SONIDO ENVOLVENTE EN LA ACTUALIDAD



Para la generación de sonido envolvente se ha recurrido a una gran cantidad de técnicas. A continuación enlistamos las más sobresalientes y exitosas:

2.1 Grabación múltiple

Esta técnica, quizá la más sencilla, está basada en la colocación de varios dispositivos de grabación, los cuales permiten recuperar la información de posicionamiento para reproducirse posteriormente. Dentro de esta técnica, podemos encontrar varias metodologías distintas. Señalaremos dos, que son las más utilizadas y de las cuales pueden derivarse otras técnicas similares:

2.1.1 Microfonía binaural

Consiste en la colocación de dos micrófonos en una cabeza real o falsa con características acústicas similares a las de una cabeza real.

Una de sus principales ventajas es la facilidad para obtener un sonido estéreo que es muy similar al sonido estéreo que escucharía una persona como si se encontrara en ese mismo lugar. Normalmente para la reproducción se busca que se haga uso de audífonos, ya que evidentemente cada canal de audio (izquierdo y derecho) fue obtenido exactamente en esa posición.

A continuación de listan algunas de las ventajas y desventajas que presenta esta técnica:



CAPÍTULO 2



Ventajas:

- Fácil de realizar
- Bajo costo
- No se requiere mucho conocimiento acústico para lograr muy buenos resultados
- No requiere post-producción ni procesamiento

Desventajas

- Al ser una técnica directa, no se puede alterar la posición virtual del escucha, pues esta está definida desde el momento de la grabación.
- Funciona solamente con sonidos obtenidos de manera natural. Cualquier sonido generado por computadora, por otro sistema, etc. no puede utilizar esta técnica y por lo tanto no se puede generar sonido envolvente a partir de sonidos artificiales con esta técnica.



Figura 2.1 'Dummy head' de la marca Neuman, producto comercial para la generación de audio tridimensional basado en esta técnica.

2.1.2 Grabación múltiple con reproducción múltiple

Similar a la técnica anterior, con 2 o más micrófonos, esta técnica pretende eliminar algunas de las desventajas de la técnica anterior, principalmente el hecho de que solamente se pueda reproducir mediante audífonos. Así, al tener información del sonido grabado en 2 o más lugares, un procesamiento es capaz de generar un mapa tridimensional donde estén ubicados los distintos objetos. Con ayuda de las mismas grabaciones, y la información obtenida de la posición de



CAPÍTULO 2



cada fuente de sonido, un segundo procesamiento mapea las grabaciones disponibles a una matriz de bocinas, enviándoles a cada una de ellas una mezcla distinta de las grabaciones, acorde con su posición relativa al escucha. Sin embargo, estas modificaciones en el procedimiento a seguir genera a su vez ventajas y desventajas:

Ventajas:

- No está limitada al uso de audífonos.
- Se pueden usar tantas bocinas como sea posible, utilizando procesamiento digital para la generación de los canales correspondientes.
- La microfonía no está limitada a una cabeza virtual. Siempre que el procesamiento lo permita, la colocación de los micrófonos es independiente del sistema auditivo humano.

Desventajas:

- Requiere una gran cantidad de procesamiento.
- Dado que el procesamiento normalmente son metodologías desarrolladas con propósitos comerciales, tiende a encarecerse debido al uso de tecnologías propias.
- Dada la complejidad de ambos procesamientos, no es posible utilizarse para aplicaciones en tiempo real. En otras palabras, las posiciones de las fuentes de sonido virtuales deberán definirse en una etapa previa a su



LA GENERACIÓN DE SONIDO ENVOLVENTE EN LA ACTUALIDAD



reproducción (normalmente, para aplicaciones comerciales tales como películas, se está haciendo referencia a la etapa de post-producción).

2.2 Principio de Huygens

La generación de sonido envolvente basado en física de ondas y en particular en el principio de Huygens para simularlo es quizá la técnica que en el largo plazo triunfará sobre las demás técnicas. Esencialmente, se postula que con una cantidad suficiente de bocinas es posible producir casi cualquier patrón de interferencia en la entrada del oído del escucha. Desafortunadamente la física de ondas es todavía muy difícil de simular, al menos en tiempo real, con la capacidad de cómputo actual. Aunque es una técnica muy prometedora, ya que es simplemente una simulación completa del fenómeno físico detrás de la percepción del sonido, hoy en día es poco práctica. Sus aplicaciones actuales quedan fuera del campo de aplicación de la técnica que se pretende desarrollar, puesto que hoy en día es difícil que mediante esta técnica se logren resultados en tiempo real.

2.3 Head Related Transfer Function

Como su nombre lo indica, Función de Transferencia Relativa a la Cabeza (en español), es una metodología muy popular en la actualidad que busca sintetizar los resultados de la propagación del sonido a través del aire y cómo éstos se reciben en el oído afectados también por la cabeza del escucha, su relación con cada uno de los órganos auriculares (izquierdo y derecho) y finalmente, los fenómenos que tienen lugar en la parte neuronal de la percepción auditiva. Así, se



CAPÍTULO 2



evita profundizar en la física del sonido, y en su lugar, esta metodología se enfoca en todas aquellas alteraciones al sonido que el cerebro utiliza como pistas para determinar la posición de una fuente de sonido. Por lo tanto, esta metodología pretende aplicar las mismas alteraciones a una fuente de sonido dada y así “engañar” al cerebro, dándole las mismas pistas que busca en una fuente de sonido real.

¿Qué información obtiene el cerebro de cada uno de los oídos, que le permite definir el origen de una fuente de sonido?

Esta es la pregunta fundamental que responde la metodología de la Head Related Transfer Function, la cual busca obtener una función de transferencia de la forma

$$H(s) = \frac{Y(s)}{X(s)},$$

donde

$H(s)$ es la función de transferencia en cuestión,

$Y(s)$ es la salida producida por la función de transferencia $H(s)$ a partir de

$X(s)$, la entrada a la función de transferencia, en este caso una fuente de sonido,

de tal manera que genere los mismos efectos por los que atravesaría una onda de sonido que llega a cada oído para fines de localización.

¿Cómo obtener y cómo aplicar esta función de transferencia? En la respuesta a esta pregunta es donde divergen las diferentes técnicas actuales, pues si bien todas buscan generar el mismo resultado, no todas coinciden en la manera de lograrlo.



LA GENERACIÓN DE SONIDO ENVOLVENTE EN LA ACTUALIDAD



Así pues, al simular las características del oído y la propagación del sonido desde un punto de vista psicoacústico (objeto de estudio de los siguientes dos capítulos), se puede optar por versiones que utilizan una completa transformación al dominio de la frecuencia, hacer el procesamiento necesario, y entonces volver al dominio del tiempo para presentar el resultado, bien aplicar la transformación directamente en el dominio del tiempo, o bien utilizar alguna otra metodología que lleve al resultado planteado.

Esta metodología de la Head Related Transfer Function tiene el potencial de que si, para una fuente de sonido $X(s)$ dada, la función de transferencia $H(s)$ produce la salida $Y(s)$ lo suficientemente rápido (veremos que se entiende por suficiente), entonces se habrá logrado un sistema en tiempo real que permita generar audio envolvente para aplicaciones interactivas, donde la respuesta del sistema es instantánea. Esta metodología será la que se utilizará en los capítulos 5 y 6, la implementación del modelo, para llegar a una técnica propia.

CAPÍTULO 3

EL OÍDO HUMANO

*Sabiduría es la recompensa que se obtiene
de por vida al escuchar cuando se hubiera
preferido hablar*
- Doug Larson



El presente capítulo tiene por objeto dar una caracterización del oído humano, en términos que nos permitan conocer sus capacidades y limitantes, para así poder dar inicio a la creación de una técnica que al ser aplicada a una fuente de sonido permita al oído definir de donde proviene dicha fuente, de manera simulada (es decir, engañando nuestra percepción).

3.1 Respuesta en frecuencia

Una de las primeras características que tenemos que conocer sobre el oído humano es su respuesta en frecuencia. Históricamente, al tratar de medirla, los experimentos mostraron que la respuesta en frecuencia del oído depende de la intensidad del sonido utilizado. Esto es de esperarse pues la gran mayoría de los sistemas en el mundo real exhiben este comportamiento.

Por ejemplo, en bajas intensidades de sonido (debajo de 40 dB NPS), la respuesta en frecuencia es muy variable, teniendo un máximo alrededor de los 4 kHz. Este máximo se ha teorizado en razón de que la voz humana tiene una gran cantidad de información en este rango de frecuencia, por lo cual se cree que el oído humano desarrolló una mayor sensibilidad para estas frecuencias en una etapa posterior de su evolución, sentando las bases para el desarrollo del lenguaje.

Sin embargo, para intensidades fuertes, por encima de los 100 dB NPS, la respuesta en frecuencia del oído es prácticamente plana.

Esta diferencia llevó a la construcción de curvas de nivel para entender la respuesta en frecuencia del oído humano tomando en cuenta las diferentes



CAPÍTULO 3



intensidades de los sonidos percibidos. El gráfico es muy utilizado en el área de la audición humana, y fue creado en la década de los años 30 del siglo XX por Fletcher y Munson. En la mayoría de las personas se obtienen curvas similares, mostradas en la figura 3.1

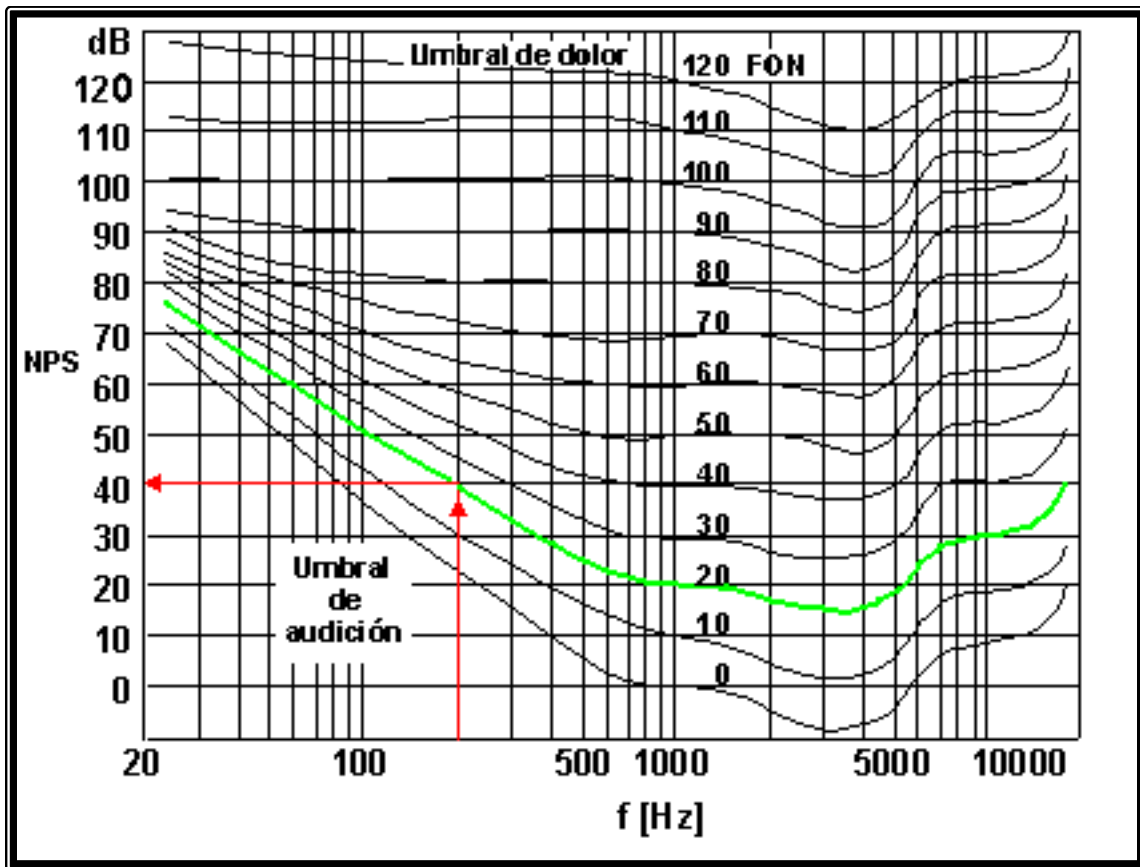


Figura 3.1 Curvas de sensibilidad del oído humano en incrementos de 10 Fones. Nótese que ambos ejes tienen una escala logarítmica.



Para poder entender el gráfico anterior, debemos definir las siguientes unidades:

NPS - Nivel de Presión de Sonido (o SPL, por sus siglas en inglés)

Es una medida relativa y logarítmica que nos indica la variación en la presión de aire generando así un sonido de mayor o menor intensidad según este valor. El valor de referencia para esta medida normalmente son $20 \mu\text{Pa}$.

Fon (o Phon, en inglés)

El fon es una unidad de intensidad subjetiva creado porque la intensidad de un sonido es percibida de manera distinta según su frecuencia. Está definido como aquel sonido que se percibe con igual intensidad a un sonido de 1 dB NPS en una frecuencia de 1 kHz.

Así, la figura 3.1 nos dice que para sonidos de baja intensidad, tomando como ejemplo la curva marcada en verde, un sonido de unos 20 Fones requiere 75 dB NPS a 20 Hz para escucharse con la misma intensidad que un sonido de 15 dB NPS a 4kHz. Dicho de otro modo, el oído tiene una ganancia de alrededor de 50 dB en los 4kHz respecto a los 20 Hz.

También debe señalarse que si bien las diferentes curvas de nivel terminan abruptamente, en realidad son curvas asintóticas con 20 Hz y 20 kHz como límites inferior y superior respectivamente. Quiere esto decir que no importa que tan intenso sea un sonido más allá de esos límites, el oído humano simplemente no lo percibirá.



3.2 Rango dinámico

La misma figura 3.1 nos indica el rango dinámico del oído humano: el umbral auditivo y el umbral del dolor. Debajo del umbral auditivo es imposible percibir un sonido. Arriba del umbral del dolor (y en intensidades cercanas a este, por períodos prolongados de tiempo) sentimos molestia en nuestros oídos en incluso podríamos perder total o parcialmente nuestra capacidad auditiva.

Siendo el rango dinámico la máxima diferencia en intensidad que se puede encontrar en la gráfica, vemos que para 4kHz el umbral auditivo está en su mínimo, de casi -10 dB SPL, mientras que para 20 Hz y 20kHz el umbral del dolor está en su máximo, de casi 130 dB, consiguiendo así el comúnmente aceptado rango dinámico de 130 dB (ya que el mínimo no es -10 dB y el máximo tampoco es 130 dB).

3.3 Patrón de captación

El patrón de captación, patrón polar o sensibilidad direccional del oído humano es la curva que nos indica la ganancia, o sensibilidad que tiene el oído con respecto a la posición en coordenadas polares de la fuente. De la misma manera como sucede con la respuesta en frecuencia, el patrón de captación depende marcadamente de la frecuencia del sonido utilizado para obtener esta curva. En un estudio detallado del oído exterior (comúnmente llamada oreja o bien, lóbulo auricular) veríamos cómo todas sus curvaturas influyen en estas gráficas. Sin embargo, para fines de desarrollar una técnica por síntesis, no es de interés estudiar esto a detalle, sino simplemente obtener los resultados experimentales



EL OÍDO HUMANO



que nos permitan igualar las gráficas mostradas en las figura 3.2, 3.3, 3.4, 3.5 y 3.6. Para estas gráficas polares, 90 grados es en la dirección frontal del escucha, siendo 180 la dirección posterior, y 0 y 180 grados la derecha e izquierda, respectivamente.

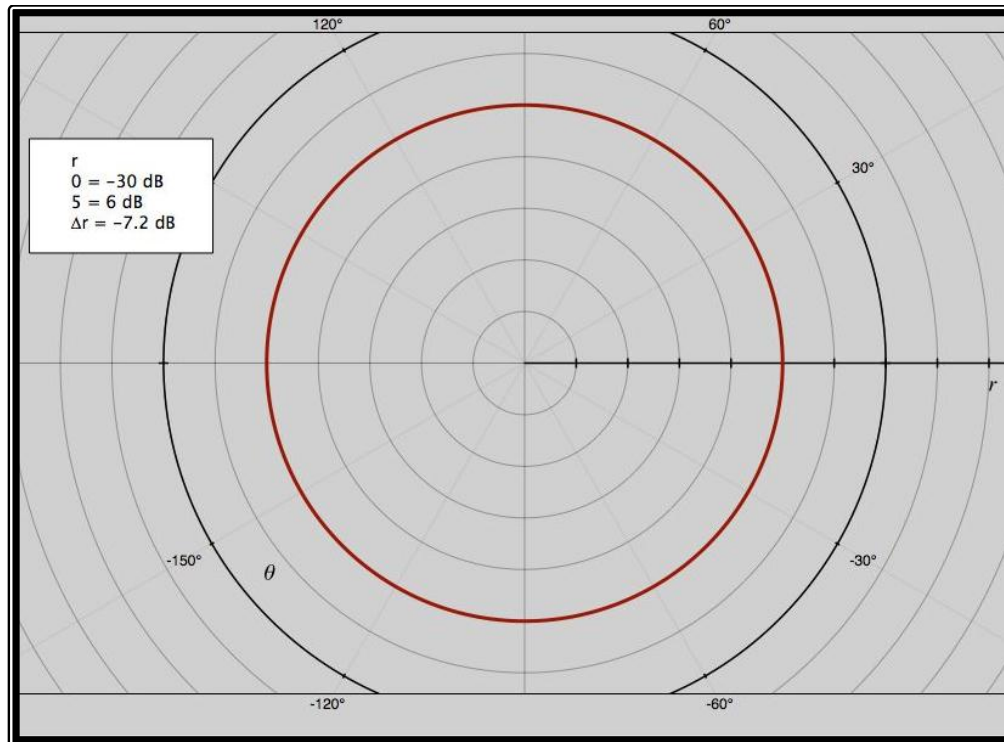


Figura 3.2 Patrón de captación del oído izquierdo, en 100 Hz (curva roja). Observamos que en esta frecuencia el patrón de captación es omnidireccional. Si no hay más pistas sobre la ubicación de la fuente, el escucha típico percibe el sonido como proveniente de adelante.



CAPÍTULO 3

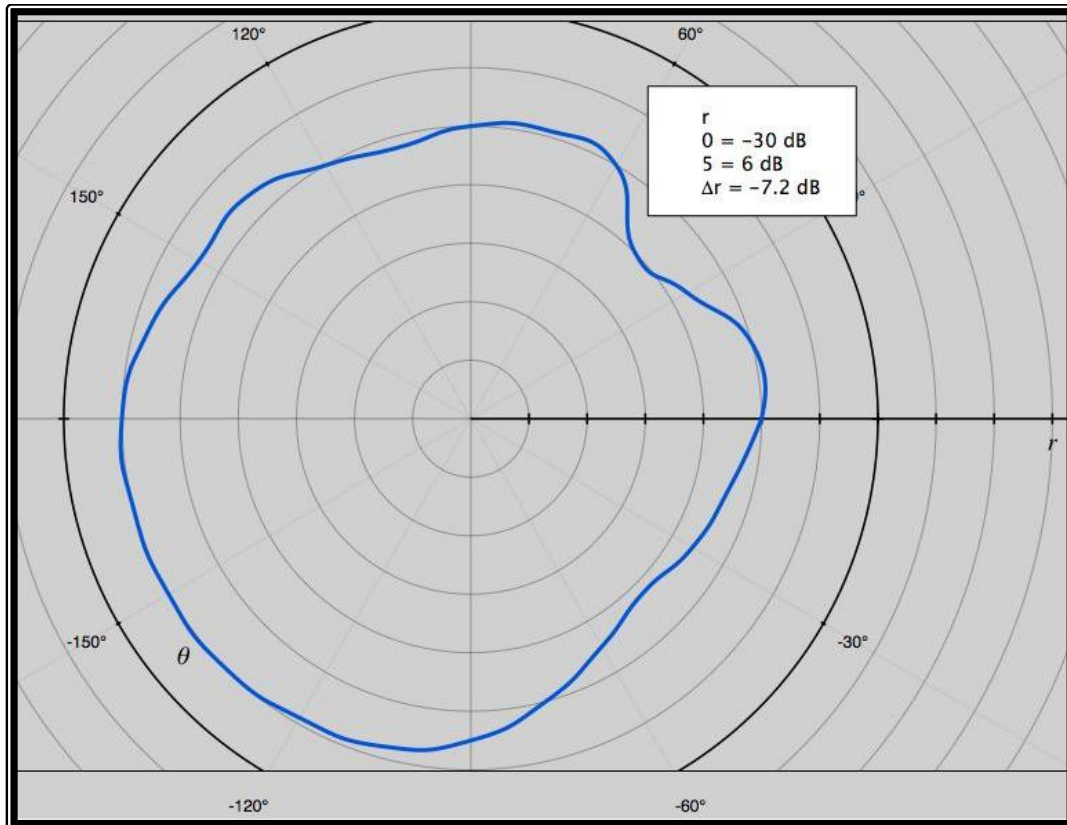


Figura 3.3 Patrón de captación a 1250 Hz. De manera similar a lo que sucede con una fuente de sonido de 100 Hz, en ausencia de más información, un sonido de 1250 Hz se percibe como proveniente desde el frente.



EL OÍDO HUMANO

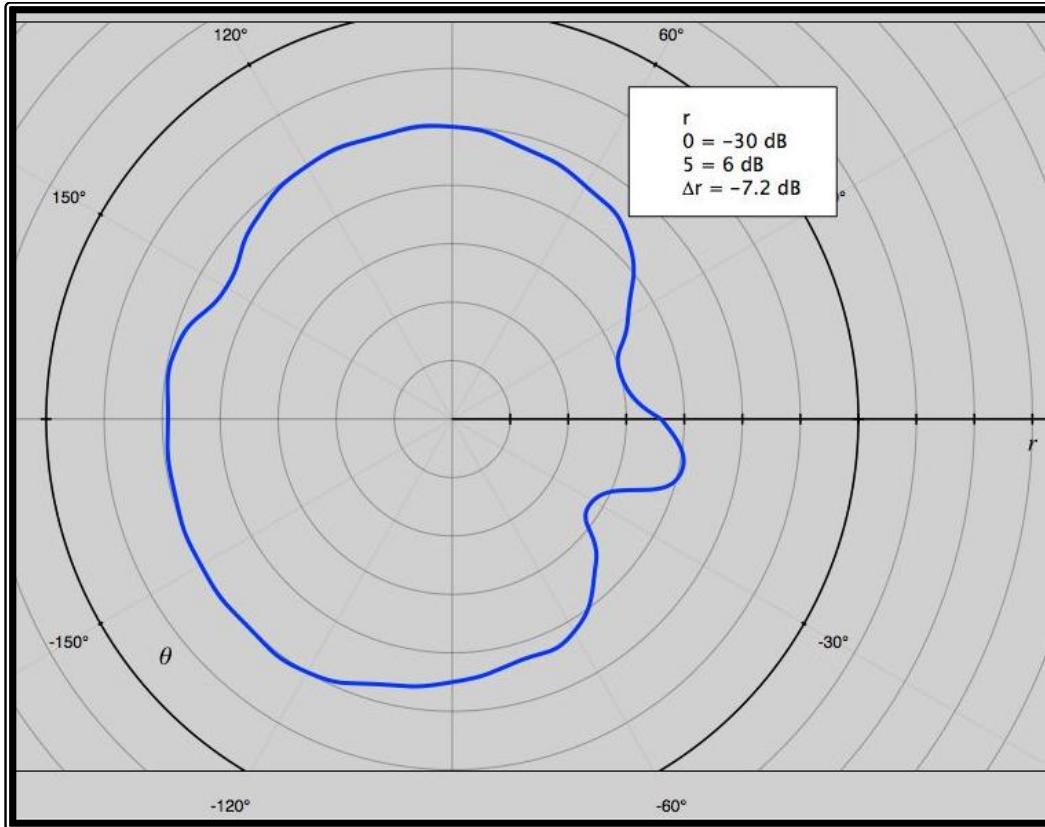


Figura 3.4 Patrón de captación usando una fuente de sonido de 2000 Hz. Nótese el aumento en la directividad.



CAPÍTULO 3

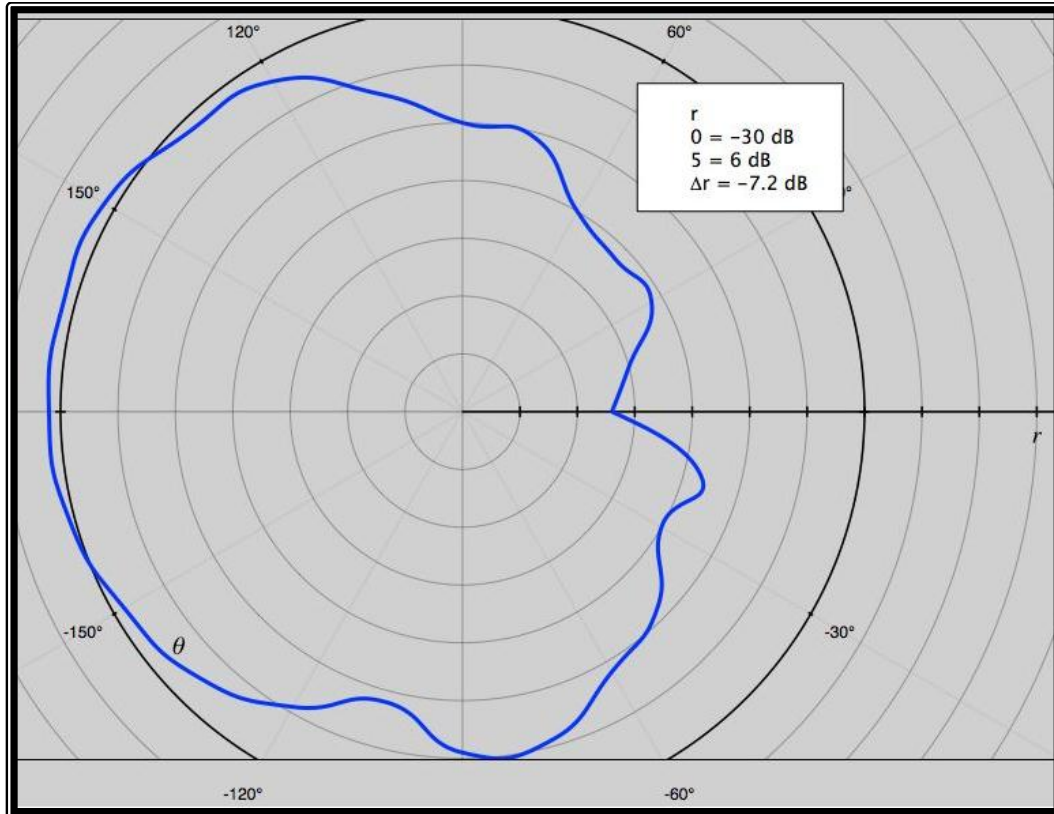


Figura 3.5 Patrón de captación usando una fuente de 5000 Hz. Nótese la elevada ganancia para los sonidos provenientes del lado izquierdo.

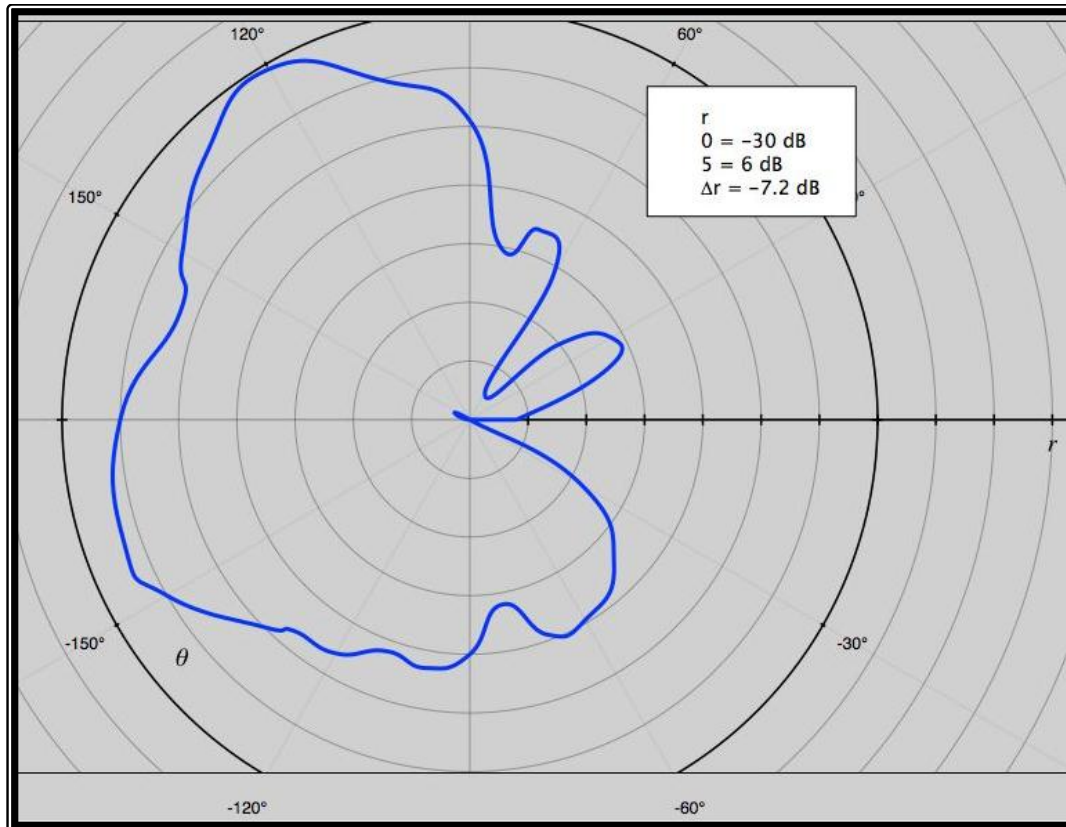


Figura 3.6 Patrón de captación con una fuente de sonido de 16 kHz. La directividad es mucho mayor en este caso.

Nótese como, por ejemplo, para una fuente de 16 kHz, un sonido que proviene del lado izquierdo tiene una ganancia relativa de 6 dB, mientras que para el oído derecho (al invertir simétricamente la figura 3.6 con respecto al eje 90 a -90) se tienen una ganancia relativa de hasta -30 dB. Es decir, para un sonido que proviene del lado izquierdo, la diferencia total en decibeles NPS entre un oído y otro es de hasta 36 dB en altas frecuencias.



CAPÍTULO 3



Esto puede contrastarse con sonidos de baja frecuencia, donde no importa la posición de la fuente, la diferencia entre un oído y otro es nula.

3.4 Fase y tiempo de llegada

Dada la separación física de nuestros dos oídos, es natural esperar que el tiempo de llegada de una fuente de sonido a un oído con respecto al otro presente una ligera diferencia. Esta diferencia es utilizada por el cerebro para crear un mapa de la ubicación de la fuente. Esto se verá con más detalle en el capítulo siguiente dado que para su estudio intervienen otros factores propios de la propagación del sonido. Asimismo, es de esperarse que no en todos los casos y para todas las frecuencias la fase sea la misma. Este fenómeno, resultado directo de las dimensiones de nuestra cabeza afectando de manera distinta cómo es que los distintos componentes de frecuencia llegan a cada oído, puede generar que un frente de onda esté en 180° en un oído, mientras que llega en 0° en el otro. Esto se verá con más detalle en el siguiente capítulo, al estudiar la propagación del sonido.

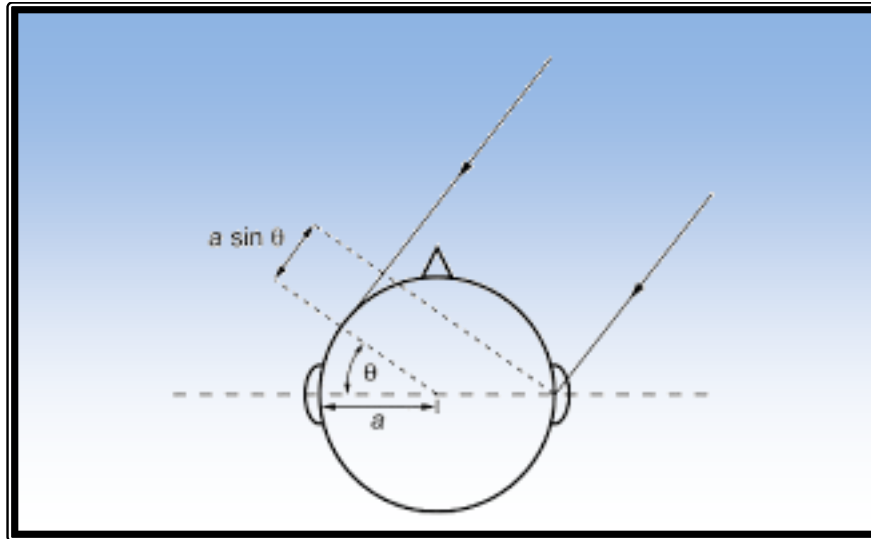


Figura 3.7 Diferencia en el tiempo de llegada de una fuente de sonido dada su velocidad limitada en el aire.

3.5 Parámetros psicoacústicos relevantes

Para fines de este trabajo, estudiaremos a continuación los parámetros del oído que son relevantes para el desarrollo de la técnica que nos permita simular sonido envolvente en tiempo real.

3.5.1 Latencia

La latencia de una señal de audio en aplicaciones comerciales es el retraso que un dispositivo, por el simple hecho de procesar la señal, le da a una señal de audio de salida medida normalmente en milisegundos con respecto a la señal de entrada.



La figura siguiente muestra el diagrama de este fenómeno:

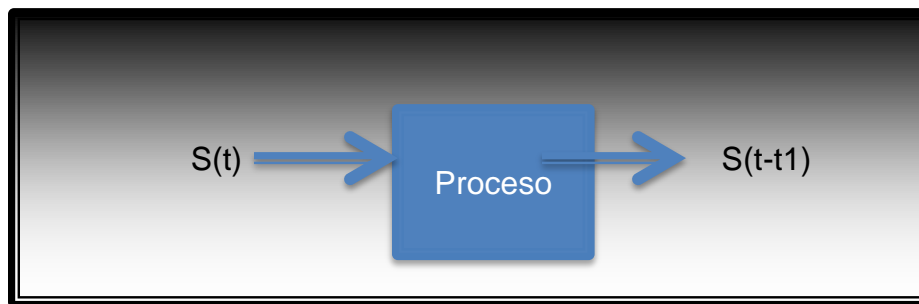


Figura 3.8 Al procesar una señal de sonido $S(t)$ se produce irremediamente un retraso, t_1 , a la señal de salida.

En aplicaciones de audio y video, es sabido que el cerebro comienza a detectar una falta de sincronía entre lo que ve y o que oye cuando el sonido tiene una latencia de más de 40 ms. En algunas situaciones, como la sincronía de labios (o lipsync, en inglés), este valor puede ser hasta de un máximo de 25 milisegundos.

Este valor impone la mayor restricción para aplicaciones de audio envolvente en tiempo real: El máximo retraso de un procesador que genere una señal de sonido envolvente a partir de un sonido dado no puede exceder este límite. Por ello, definiremos como procesamiento en tiempo real a todo aquel proceso que tenga una duración total máxima de 25 o 40 ms, según sea la aplicación práctica (el límite de 25 ms tiende a ser más aplicable a la falta de sincronía entre el video y el audio de una persona hablando. Para los demás casos, 40 ms suele ser el límite aceptado).



3.5.2 Efecto Franssen

Dada la baja sensibilidad de los oídos a cambios en la fase de una fuente de sonido, se ha demostrado experimentalmente que los primeros instantes de la percepción son los más importantes para su ubicación. Una vez localizado el sonido, si este tiene pocos cambios en amplitud o en su espectro de frecuencia, puede incluso llegar a desplazarse a una posición final distinta de la inicial sin que un escucha perciba dicho movimiento. Esto es importante porque al optimizar una técnica de procesamiento de esta naturaleza, es posible generar pistas de ubicación sólo en los primeros instantes, y después eliminar cualquier procesamiento adicional, aligerando la carga de procesamiento digital de un sistema dado sin comprometer los resultados deseados.

CAPÍTULO 4

PROPAGACIÓN DEL SONIDO

El barco vacío es el que más sonido produce
- William Shakespeare



En este breve capítulo mostraremos los parámetros importantes del sonido para fines de este trabajo.

4.1 El sonido como una onda

El sonido es el resultado del movimiento de las partículas de aire al ser afectadas por un objeto en movimiento, normalmente en algún tipo de vibración. Este movimiento en las partículas de aire causa una diferencia de presión (al desplazarse una cantidad significativa de partículas de un lado a otro). Esas partículas, a su vez, empujan a otras partículas, cediendo ese aumento de presión a otra región y regresando a su estado original. Este proceso se repite sucesivamente, perdiéndose energía paulatinamente hasta que la presión se vuelve a estabilizar. De esta breve explicación surgen varias preguntas, que se responderán en los siguientes apartados: ¿A qué velocidad se propaga ese cambio de presión? ¿En qué dirección se propaga dicho cambio de presión? ¿Hay alguna relación entre la frecuencia de la vibración y la velocidad a la que se propaga la diferencia de presión?

El sonido no es otra cosa que una onda, fenómeno ampliamente estudiado en la Física. Sin embargo, la importancia del sonido en la vida del hombre es tal, que existe una rama de la física dedicada a estudiar, de entre todas las ondas, al sonido en particular, ya que de su entendimiento se desprenden muchas aplicaciones en la actualidad: equipos de sonido, aparatos de ayuda auditiva, diseño de salas de concierto, etc. Dicha rama de la Física se conoce con el nombre de Acústica. A continuación se muestran varias características del sonido,



CAPÍTULO 4



desde el punto de vista de la Acústica, que son relevantes para su interacción con el oído y el posterior desarrollo de la técnica de generación de audio envolvente.

4.2 Frecuencia

Si bien la Acústica estudia el sonido como una onda mecánica cuyo espectro de frecuencia no está limitado por el oído humano, aquí se tratará exclusivamente las ondas sonoras que estén en el rango de 20 Hz a 20 kHz, que es el rango aceptado de frecuencias que se pueden percibir por el oído humano.

4.3 Velocidad de propagación

El sonido, siendo una onda mecánica cuyo principal medio es el aire, tiene una velocidad de propagación media de 343 m/s. Velocidades diferentes aplican para otros medios como sólidos (maderas, metales) o líquidos (como agua). En este trabajo se estudiará principalmente al sonido como una onda que viaja por el aire, y en algunos capítulos se mencionará su propagación en algunos otros medios.

En el aire, hay muchos factores que determinan el valor exacto de esta velocidad de propagación, como la temperatura, la presión atmosférica, la humedad, por lo que solamente podemos considerar este como un valor promedio. Esta cantidad se usará en el resto de este trabajo como un valor que permitirá dar una idea de las magnitudes envueltas, pero no debe de perderse de vista que esta velocidad no siempre es la misma.



También es muy importante mencionar que como cualquier otra onda mecánica, el sonido se puede propagar por cualquier medio sólido, líquido o gaseoso. Para fines de este trabajo, a menos que se especifique explícitamente lo contrario, se deberá suponer que se está haciendo referencia a la propagación del sonido en el aire.

4.4 Dirección de propagación

El sonido es una onda longitudinal. Esto significa que la dirección en la cual se producen las vibraciones es paralela a la dirección en la cual se produce la propagación de la onda. Este concepto será utilizado más adelante para revisar el fenómeno de refracción del sonido.

4.5 Longitud de onda

La longitud de onda del sonido está dada por la física de ondas, en particular, por la siguiente ecuación:

$$\lambda = \frac{v}{f}$$

donde:

λ = longitud de onda, en metros,

v = velocidad del sonido, 343 m/s,

f = frecuencia del sonido cuya longitud de onda deseamos conocer, en Hz



CAPÍTULO 4



De tal manera que para $f = 20$ Hz, la correspondiente longitud de onda será de poco más de 17 m, mientras que para 20 kHz, se corresponde una longitud de onda de 1.715 cm.

La importancia de esto radica en que, como se verá a detalle más adelante, aunque los mismos fenómenos regidos por las mismas leyes aplican a cualquier sonido, los resultados dependen de las magnitudes particulares de la onda estudiada. Es decir, los fenómenos que se revisen a continuación ofrecerán diferentes resultados según la frecuencia del sonido utilizada.

4.6 Frente de onda

El frente de onda, definido como el lugar geométrico de los puntos que corresponden a una misma fase, para un instante dado, de una onda en propagación, es un concepto bastante útil que sirve, entre otras cosas, para visualizar a las ondas de una manera más entendible en el estudio de éstas, sean acústicas, electromagnéticas, etc.

El concepto de frente de onda nos ayudará a estudiar en este trabajo el fenómeno de refracción.

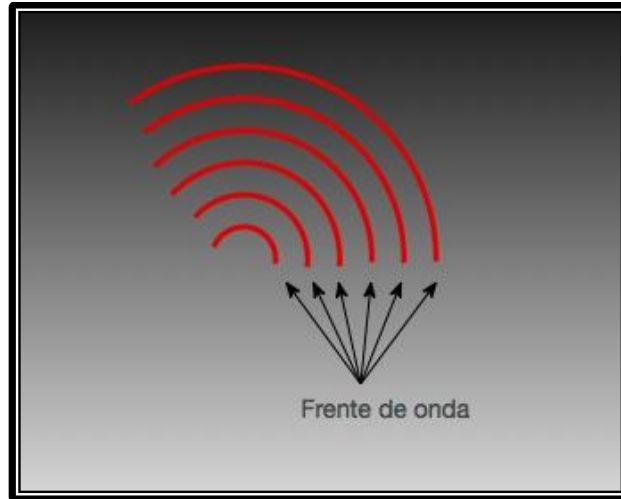


Figura 4.1 Diagrama típico de un frente de onda mostrando curvas donde la onda presenta la misma fase

4.7 Refracción

La refracción del sonido sucede cuando este se encuentra en su camino (conforme se propaga) con un medio de transmisión de características diferentes al medio en donde se encuentra.

Esencialmente, se produce un cambio en la dirección del frente de onda cuando la onda en propagación del medio físico 1 llega al medio físico 2 de manera no perpendicular, en cuyo caso normalmente no se presenta un cambio. Este cambio en la dirección es importante para este trabajo ya que, como veremos en el capítulo 5, la cabeza de una persona refracta el sonido de tal forma que influye tanto en tiempo de llegada como en la intensidad del sonido en un oído con respecto al otro, debido a que altera la propagación del sonido. En este capítulo



CAPÍTULO 4



sólo se menciona la refracción como un concepto, y en el capítulo siguiente se detallará como funciona al estudiar específicamente la propagación de sonidos interactuando con el oído humano y la cabeza como objeto refractor.

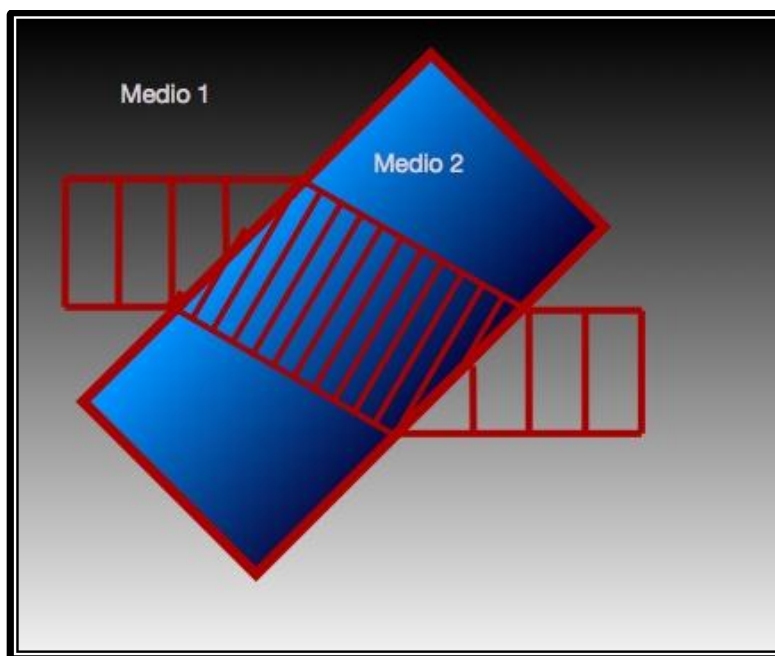


Figura 4.2 Cambio en la dirección de un frente de onda al pasar del medio 1 al medio 2 y después, del medio 2 al medio 1

Es importante mencionar que la frecuencia de la onda propagada no se ve alterada. Sin embargo, dado el cambio en las características del medio físico 2 con respecto al medio físico 1, la longitud de onda se ve modificada debido a que la velocidad de propagación de la onda es distinta. No hace falta detallar la teoría detrás de este fenómeno, ya que para el desarrollo de la técnica de procesamiento de audio a generar basta con tomar en cuenta los efectos prácticos de la refracción. Además, hablar a detalle del fenómeno de refracción de ondas bien



podría llevarse un capítulo por sí mismo, sin tener mucha relevancia para fines de este trabajo.

4.8 Difracción

La difracción es un fenómeno que se da cuando una onda en propagación interactúa con un objeto de dimensiones cercanas a la longitud de onda de la onda en cuestión, el cual no sólo puede llegar a alterar la dirección de propagación de la onda sino también la forma del frente de onda.

Es ampliamente estudiado tanto en las disciplinas de Física de Ondas así como en Acústica. Se puede explicar de manera sencilla a partir del principio de Huygens, el cual establece que cualquier punto de un frente de onda puede considerarse como una fuente esférica de una onda con la misma frecuencia que la onda original, a su vez que dicho punto también es la suma de todos las fuentes esféricas de todos los puntos anteriores, considerados de igual manera como fuentes esféricas. Frentes de onda posteriores en el tiempo (conforme la onda se propaga) pueden predecirse *a priori* partiendo de que dichos frentes de onda serán la suma de las fuentes esféricas de los puntos actuales, facilitando así la predicción de los patrones de interferencia o frentes de onda que se generan cuando una onda interactúa con objetos.

Este fenómeno explica por qué, en la experiencia cotidiana, un sonido parece provenir desde la entrada de una habitación, pues es la entrada considerada una



fuentes de sonido, según el principio de Huygens, de un sonido que realmente proviene de algún lugar dentro de dicha habitación.

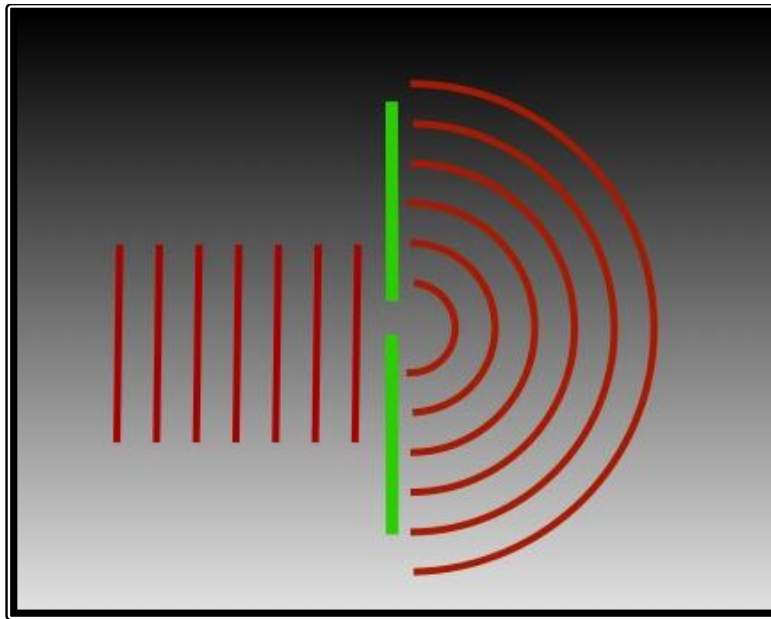


Figura 4.3 Frentes de onda (rojo) mostrando el fenómeno de difracción al interactuar con la pared marcada en verde con un hueco al centro

En el capítulo 5 se verá cómo la cabeza del ser humano produce distintos patrones de propagación de una onda de sonido, según sus dimensiones.

4.9 Patrón de radiación

De la misma manera en que el oído tiene un patrón de captación, toda fuente de sonido tiene un patrón de radiación, el cual depende de la disposición física de los elementos vibrantes que generan el sonido. Esto significa que de manera general,



la directividad de la fuente de sonido provocará que la ganancia no sea la misma al llegar a ambos oídos:

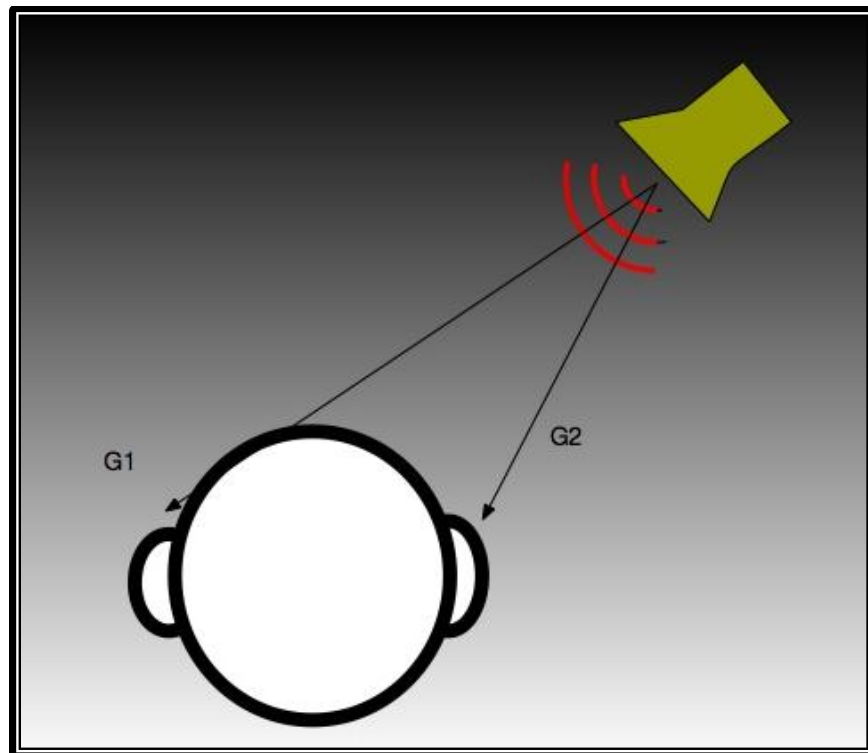


Figura 4.4 La ganancia es distinta en cada oído según el patrón de radiación de la fuente de sonido

En la mayoría de las situaciones, G_1 , la intensidad del sonido en el oído izquierdo será diferente de la intensidad en el oído derecho, G_2 . Pero vale la pena analizar un poco esta diferencia en ganancia: el cerebro es capaz de identificar la posición de fuentes de sonido que jamás ha escuchado. Si jamás las ha escuchado, el patrón de radiación debe ser desconocido en la experiencia del individuo que



CAPÍTULO 4



percibe dicha fuente. ¿Qué quiere esto decir? Que en la práctica, esta diferencia de ganancia es poco relevante para fines de ubicar la posición de la fuente.

Lo anteriormente dicho quiere decir que el cerebro se basa en información que ya conoce (el patrón de captación del oído humano, a través de la experiencia) que en información nueva que tendría que asimilar después de experimentarla por un tiempo (el patrón de radiación de la fuente de sonido). Hay que recordar que el oído, como se mencionó en el capítulo 1, evolucionó originalmente como un medio de defensa. Esto significa que como sentido, debe depender más de la experiencia del individuo que del medio ambiente para funcionar adecuadamente.

4.10 Dispersión acústica

Este fenómeno se refiere a la diferencia en la velocidad de propagación de los distintos componentes en frecuencia de un sonido, debido a que el medio propaga con mayor velocidad ciertas frecuencias con respecto a otras, conocido éste concepto como velocidad de fase.

Este fenómeno significa que el oído puede ser capaz de distinguir la distancia de una fuente de sonido basado en la dispersión que presente el sonido al llegar.

CAPÍTULO 5

EL SISTEMA OÍDO-AIRE-FUENTE DE SONIDO

*La armonía potencializa cosas pequeñas.
Su ausencia destruye cosas grandes
- Salustio*



CAPÍTULO 5



Este capítulo tiene como objetivo combinar la información recabada en los capítulos 3 y 4 sentando así las bases para modelar el sistema fuente de sonido-aire-oído, para así poder crear una técnica de procesamiento de audio digital.

5.1 Consideraciones

Si bien la propagación del sonido es un fenómeno que depende de muchísimos factores ambientales, en este trabajo se considerarán constantes, tales como la presión atmosférica, la densidad media del aire, la temperatura ambiente, humedad, viento nulo. Esto simplifica de manera general el diseño de una técnica de simulación de sonido envolvente sin disminuir notablemente la percepción de la posición de la fuente de sonido.

Adicional a esto, tampoco se tomará en cuenta la acústica del lugar. Dado el efecto Franssen explicado en el capítulo 3, cualquier efecto de ambientación o reverberación no afecta la ubicación del sonido, por lo tanto no es necesario lidiar con este aspecto.

5.2 Diferencia de tiempo interaural

El hecho de que la velocidad del sonido tenga un límite implica que un sonido arribará en distintos instantes para cada oído de un escucha. Si se considera una distancia media entre un oído y otro de 20 cm (consideración hecha el resto de este trabajo), obtendremos una diferencia de



EL SISTEMA OÍDO-AIRE-FUENTE DE SONIDO



$$tiempo = distancia/velocidad = (20\text{ cm})/(343\text{ m/s}) \approx 6\text{ ms}$$

Cabe mencionar que esto es verdad considerando que el sonido proviene desde cualquier lado del escucha, teniendo que recorrer dicha distancia entre ambos oídos en la dirección de propagación del sonido. Así también se considera que la cabeza no interviene en la propagación del sonido, como si los oídos estuvieran suspendidos en el aire.

La fórmula general para obtener este retraso se puede obtener a partir de la siguiente figura:

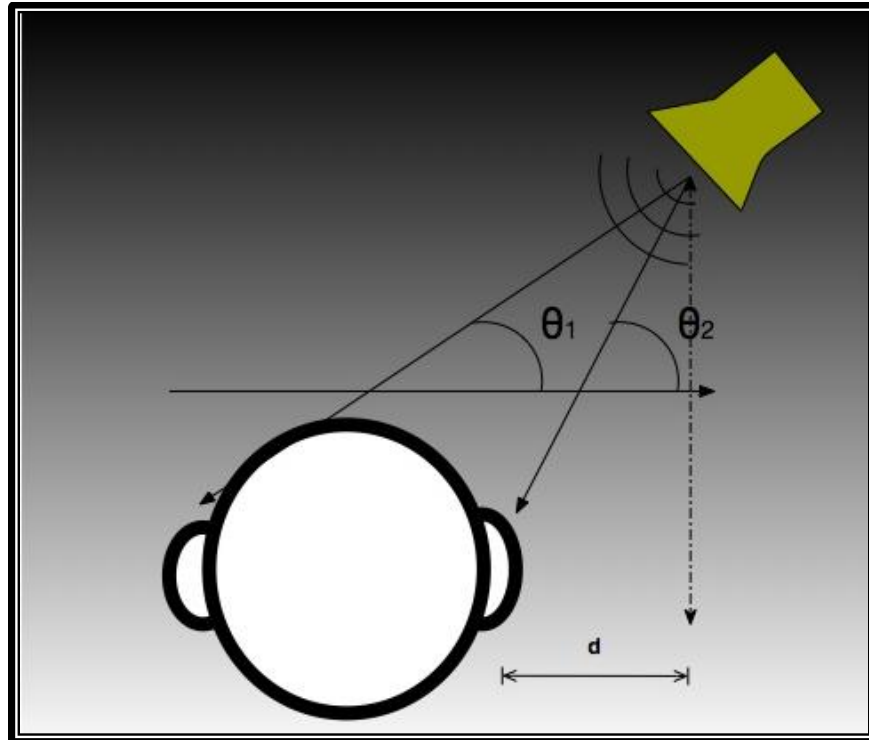


Figura 5.1 Esquema de diferencia de tiempo interaural debido a la diferencia en la distancia recorrida por la onda de la fuente de sonido para llegar a cada oído

Si se considera a cada línea como el vector velocidad de la onda de sonido en la dirección que sigue hacia cada oído, es posible obtener la componente horizontal de cada uno de los dos vectores a partir de los ángulos θ_1 y θ_2 , y así obtener el tiempo de llegada para cada oído y por ende, el retraso de un oído con respecto al otro:

$$\Delta t = t_1 - t_2 = \frac{d + 20\text{cm}}{V_{\text{sonido}} \cos \theta_1} - \frac{d}{V_{\text{sonido}} \cos \theta_2}$$



EL SISTEMA OÍDO-AIRE-FUENTE DE SONIDO



Es importante mencionar que para este fenómeno aislado se puede considerar que existe simetría respecto al plano sagital y al plano coronal, planos utilizados en la anatomía humana y mostrados a continuación:

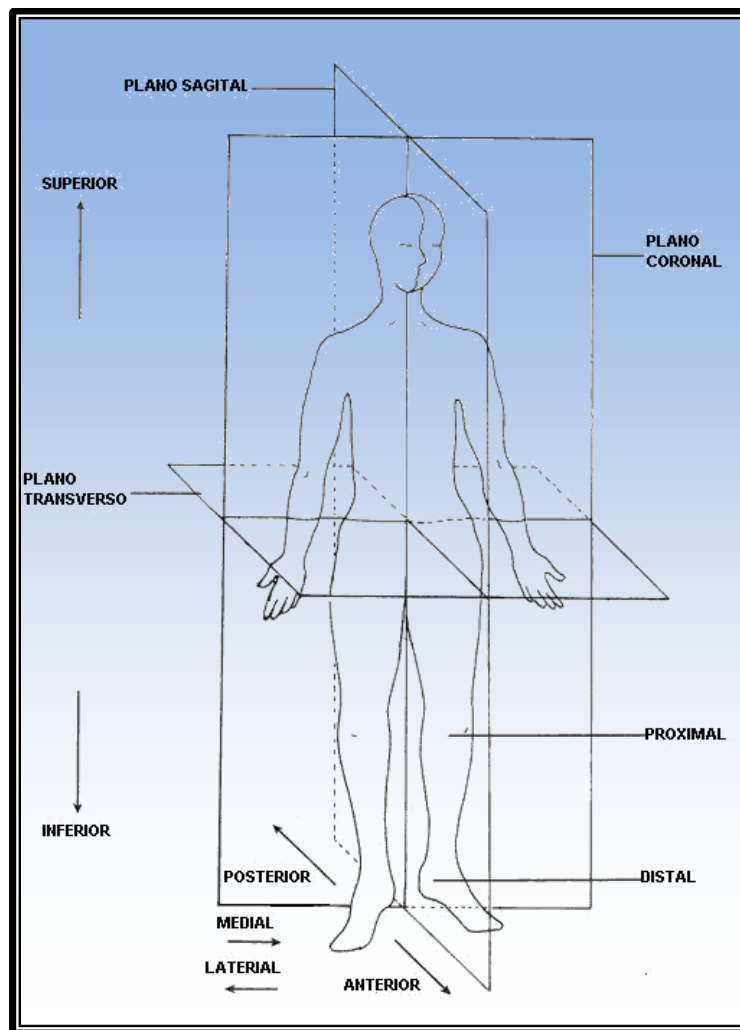


Figura 5.2 Planos anatómicos



CAPÍTULO 5



Por lo tanto es suficiente con conocer este fenómeno en el primer cuadrante trigonométrico para así usar la simetría respecto a estos dos planos y conocer el fenómeno en el resto de los cuadrantes.

También vale la pena mencionar que para distancias grandes de la fuente de sonido, se puede considerar a los ángulos θ_1 y θ_2 como iguales, misma simplificación que sucede cuando el objeto está a nuestra derecha o izquierda, siendo estos ángulos iguales entre sí e iguales a 0:

$$\Delta t = t_1 - t_2 = \frac{20cm}{V_{sonido}}$$

Que es la misma fórmula que se obtuvo al principio de esta sección.

5.3 Fase de llegada, dirección de propagación y retraso de tiempo adicional

En el apartado anterior se hizo la mención de que la ecuación de diferencia de tiempo interaural desarrollada no considera los efectos acústicos de la cabeza en la onda de sonido. Sin embargo, en la realidad, la cabeza tiene un papel muy importante para generar pistas de localización del sonido. Su presencia puede atenuar algunas frecuencias con mayor o menor medida, producir diferencias de tiempo mayores o menores según la componente de frecuencia analizada. La siguiente figura ilustra cómo una cierta longitud de onda de tamaño comparable a



las dimensiones de la cabeza modifica su frente de onda por la simple presencia de la cabeza:

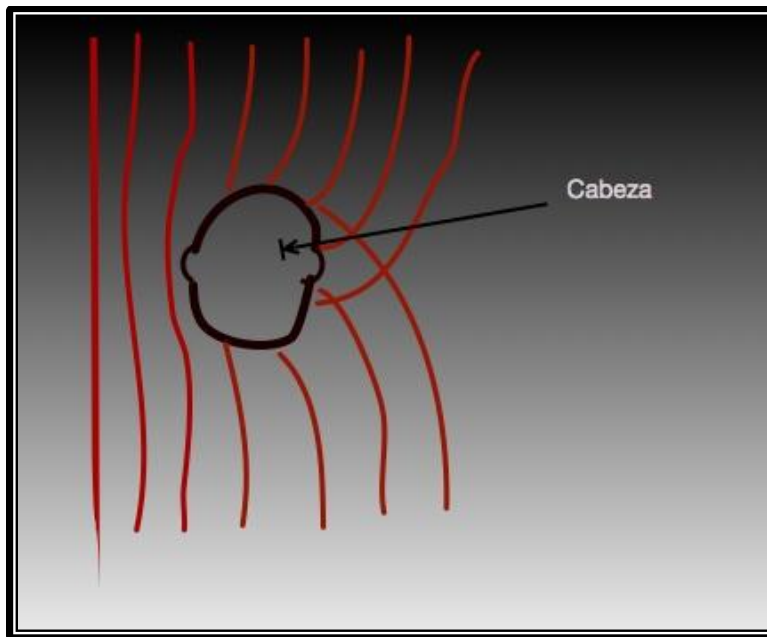


Figura 5.3 Diagrama mostrando el cambio en la forma del frente de onda por la presencia de la cabeza. Obsérvese como el comportamiento es distinto para cada oído.

Todas estas diferencias pueden obtenerse de manera teórica mediante un modelo de simulación acústica completo, el cual tome en cuenta todas las variables envueltas, tales como la densidad media de la cabeza, del aire, la forma del oído externo, la dispersión acústica, las leyes de la refracción y difracción, etc. Sin embargo, no es el propósito desarrollar una simulación exhaustiva de este sistema, sino obtener un modelo que genere al menos los mismos efectos relevantes que el cerebro utiliza para determinar la posición de una fuente de sonido. Por lo tanto, la mayoría de estos valores se obtendrán en el capítulo 6



CAPÍTULO 5



mediante experimentación. Aun así es importante conocer los fenómenos subyacentes, pues permite acotar el universo de modificaciones que habrá que hacer a una señal de sonido para generar el mismo efecto.

5.4 Función de transferencia

Como se vio en el capítulo 4, los distintos fenómenos que rigen la física de ondas, en particular los estudiados en los apartados 4.3: velocidad de propagación, 4.7: refracción, 4.8: difracción y 4.10: dispersión acústica, producen en su conjunto lo que se denomina como la función de transferencia del sistema.

Esta función de transferencia es la que pretende igualarse en términos de su salida a partir de su entrada, sin igualar internamente los pasos que sigue la función de transferencia para producir estos cambios. (Vale la pena recordar que en el capítulo 2 se estudiaron las distintas metodologías para la generación de audio envolvente, y se dijo que la metodología a utilizar en este trabajo sería la de sintetizar la función de transferencia, sin profundizar en los fenómenos físicos).

5.5 Modelado del sistema

De manera simplificada el sistema que se simulará es el siguiente:



EL SISTEMA OÍDO-AIRE-FUENTE DE SONIDO

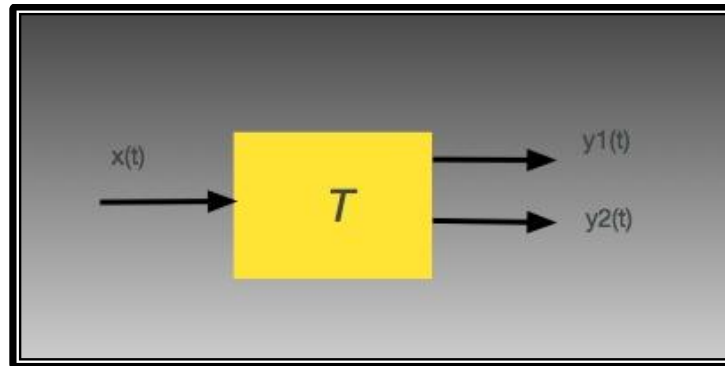


Figura 5.4 Diagrama simplificado del sistema a simular. Una entrada y dos salidas

Como puede observarse en la figura 5.4, el sistema T generará $y1(t)$ y $y2(t)$ a partir de $x(t)$, el sonido original. Dicho sistema debe producir efectos equivalentes para fines de localización, los cuales fueron revisados en los capítulos 3, 4 y 5.

Vale la pena mencionar que este es un sistema lineal e invariante en el tiempo:

- Lineal porque dadas dos fuentes de sonido introducidas al sistema (combinadas como una sola entrada, la suma de ambas), el cerebro percibe su ubicación de manera independiente una de otra. Es decir:

$$\text{si } x(t) = x_a(t) + x_b(t), \text{ entonces } T(x(t)) = T(x_a(t)) + T(x_b(t))$$

- Invariante en el tiempo porque, como se dijo al principio de este capítulo, todos los parámetros acústicos variables en el tiempo, tales como la presión atmosférica, densidad del aire, etc. son considerados constantes, por lo cual no hay dependencia de este sistema con respecto al tiempo. Si se



CAPÍTULO 5



habla de una implementación digital del sistema, bien podría decirse invariante a desplazamientos, pues es el término equivalente para sistemas discretos.

CAPÍTULO 6

IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL

*La creatividad requiere el valor
de abandonar la certidumbre*
- Erich Fromm



CAPÍTULO 6



Ahora que ya se cuenta con el modelo, este capítulo busca la manera de llevarlo a la realidad haciendo uso de las herramientas que se encuentran disponibles de manera pública.

6.1 Elección de las tecnologías disponibles

Este apartado tiene por objeto explorar las diferentes posibilidades a seguir para finalmente generar sonido envolvente. Existen diferentes productos comercialmente disponibles, tanto en software como en hardware, lenguajes de programación, plataformas de desarrollo, etc. Cada uno de los puntos siguientes tiene como objeto trazar los medios por los cuales estará sustentada la implementación práctica.

6.1.1 Digital o analógico

Si bien es factible llevar a cabo la implementación del sistema propuesto en un sistema analógico, los cuales normalmente son los más eficientes dado que son de propósito específico, el tiempo de diseño así como su construcción implicaría una cantidad de esfuerzo mucho mayor invertida que si se utilizara una solución digital. Esto es porque el procesamiento requerido es bajo y el equipo de cómputo disponible actual es lo suficientemente poderoso como para llevar a cabo el trabajo requerido en este caso.



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



6.1.2 Formatos de audio

Una vez decidido que se trabajará con sistemas digitales, es necesario definir el formato de audio que se utilizará como ejemplo para la implementación práctica. Como ha sido visible, las alteraciones que se la harán a una señal de sonido dada pertenecen al dominio del tiempo, salvo por la etapa de predescomposición, la cual pertenece al dominio de la frecuencia. Dada esta situación es preferible elegir con anticipación el tipo de audio que utilizaremos para ejemplificar la técnica propuesta: uno que represente al sonido en el tiempo. Además, dado que se manipulará directamente la intensidad de cada espectro de frecuencia, es preferible poder manipular la señal en términos de su amplitud.

Eso significa que para el formato de presentación de la señal (codificación), Pulse Code Modulation (PCM) es el más apto para nuestros propósitos, ya que permite modificar directamente el valor de la señal de sonido en el instante que se desee. Por otro lado, dado que se está experimentando con el rango completo de frecuencias percibidas por el oído, el formato de audio debe permitir la manipulación directa desde los 20 Hz hasta los 20 kHz.

Reducido así el universo de posibles formatos de audio, el WAV en formato PCM de 44.1 kHz es el elegido para trabajar, porque además es de uso libre y relativamente sencillo de usar. Recordemos que por el teorema de Nyquist,



6.1.3 Lenguajes de programación

En esta implementación se busca también demostrar que es posible generar una librería que permita integrarse a componentes más complejos razón por la cual se busca utilizar un lenguaje de propósito general que sea ampliamente usado tanto en dispositivos embebidos, móviles, portátiles y de escritorio. Para este caso se usará el lenguaje C, por varias razones, a continuación enlistadas:

- Alta portabilidad. C es un lenguaje bastante independiente de la plataforma objetivo, tanto del procesador como del sistema operativo. Así que el código generado será de manera general llevable a una amplia gama de dispositivos.
- Tiene un rendimiento óptimo para aplicaciones en tiempo real. Dado que normalmente en lenguaje C se tiene acceso directo a los dispositivos de entrada y salida, será más fácil cumplir con la meta de ser capaces de ejecutar una aplicación por debajo de 40 ms.
- Existe una gran cantidad de librerías actualmente que permitirán y facilitarán la manipulación de archivos y dispositivos de entrada/salida.

6.1.4 Librerías para manipulación de audio

Si bien trabajar con archivos WAV es relativamente fácil, en la práctica, lidiar con los metadatos del archivo puede ser un poco tardado para fines de esta implementación. Por ello una búsqueda rápida de las librerías que nos permiten manipular archivos WAV nos lleva a la librería llamada *libsndfile*.



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



Esta librería está escrita también en C y mediante sencillas funciones podemos comenzar a manipular archivos WAV sin mucho esfuerzo.

Para instalarla, basta hacer lo siguiente:

```
wget http://www.mega-nerd.com/libsndfile/files/libsndfile-1.0.25.tar.gz
tar -zxvf libsndfile-1.0.25.tar.gz
cd libsndfile-1.0.25
./configure
make
make install
```

Con esos sencillos pasos se habrá compilado, enlazado y colocado en la ruta pertinente a nuestro sistema la librería que nos permitirá trabajar con archivos WAV.

6.1.5 Descripción del sistema

Con todos los fenómenos auditivos estudiados en el capítulo 3 y la propagación del sonido, estudiada en el capítulo 4, se llegó en el capítulo 5 a la síntesis de lo que cualitativamente debe ser capaz de realizar nuestra técnica de procesamiento. Con el modelado físico del fenómeno de percepción auditiva, haremos ahora un sistema que funcionalmente produzca los mismos efectos sobre una fuente de sonido dada, sin que necesariamente trabaje igual de manera interna. Antes de eso, definiremos algunos conceptos que se utilizarán en el resto de este capítulo



para poder así dar una descripción formal de la técnica de procesamiento de audio:

Señal original/de entrada al sistema

Si bien el sistema final tendrá más de una entrada, por señal original o señal de entrada nos referiremos exclusivamente a la señal de audio a la cual deseamos aplicarle la técnica de procesamiento para generar audio envolvente.

Señal estéreo izquierda/Señal estéreo derecha o bien, señales de salida

Con este nombre se hará referencia a las señales que se esperan obtener después del procesamiento. Cada una ha de reproducirse en su correspondiente auricular para así generar el efecto deseado. Además, durante etapas intermedias del sistema, también se hará referencia con el mismo nombre a las señales generadas parcialmente, en la medida en que sea posible hacer referencia a los distintos caminos (rastrear) que sigue la señal original en el sistema para dar como resultado estas dos señales.

Tabla paramétrica

Esta tabla, la cual es una entrada al sistema, contiene distintos valores de ganancia, retraso e inversión para distintas bandas de frecuencia. Estos valores están asociados a una posición dada del espacio. Esta tabla puede entenderse como una función de un sistema de coordenadas espaciales x , y , z , donde a cada punto del espacio le corresponden a su vez, n -dos tuplas: la primera asociada al oído izquierdo y la segunda al derecho, siendo cada tríada de valor los siguientes:



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



ganancia, retraso e inversión (ya que la velocidad de fase puede ser opuesta, la señal estaría invertida con respecto al otro oído). El número n se refiere al número de bandas en las cuales se haya descompuesto la señal original.

Coordenadas de la fuente de sonido

Sin importar el sistema de coordenadas utilizado, la implementación usará esta información como entrada para seleccionar los parámetros adecuados de la tabla paramétrica, aplicándolos a la señal de entrada y produciendo así las salidas, que el cerebro interpretará como provenientes de la posición dada.

Es importante tomar en cuenta que a partir de una señal de sonido monaural se generarán dos señales, una para cada oído con los cambios siguientes:

- Producir diferencias en el tiempo, respecto a la señal original, de los tiempos de llegada de las señales, diferentes para cada señal de audio producida. Dichos retrasos son distintos para distintas bandas de frecuencia, obedeciendo a la diferencia en velocidad de fase relativa a la propagación del sonido, revisada en el capítulo 4, así como a posibles efectos generados por la presencia de la cabeza, afectando los tiempos de llegada de manera distinta según la frecuencia del sonido.
- De la misma manera, alterar la ganancia según la posición virtual relativa de la fuente de sonido con respecto al escucha. Como vimos en el capítulo 3, el oído tiene un patrón de captación que es altamente dependiente de la dirección de llegada de la fuente de sonido, de la frecuencia de ésta y de la



CAPÍTULO 6



distancia de la fuente (dispersión acústica). Por tanto, es necesario modificar la ganancia según estas variables.

- Como se vio en el capítulo 5, respecto a la propagación del sonido, puede suceder que, debido a la dirección de propagación del frente de onda de la fuente de sonido, la fase de llegada de un oído con respecto al otro esté completamente invertida. Esto es, mientras en un oído el sonido está comprimiendo el aire, en el otro oído se está expandiendo.

6.2 Diagrama de bloques del sistema

Para generar el efecto deseado, se hará una predescomposición de la señal de entrada en bandas de frecuencia. Este proceso puede no ser en tiempo real debido al tiempo de filtrado de toda una señal de audio. La predescomposición en bandas de frecuencia obedece a que todo el sistema a modelar genera diferentes resultados según la banda de frecuencias del sonido.



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



ETAPA DE PREPROCESAMIENTO

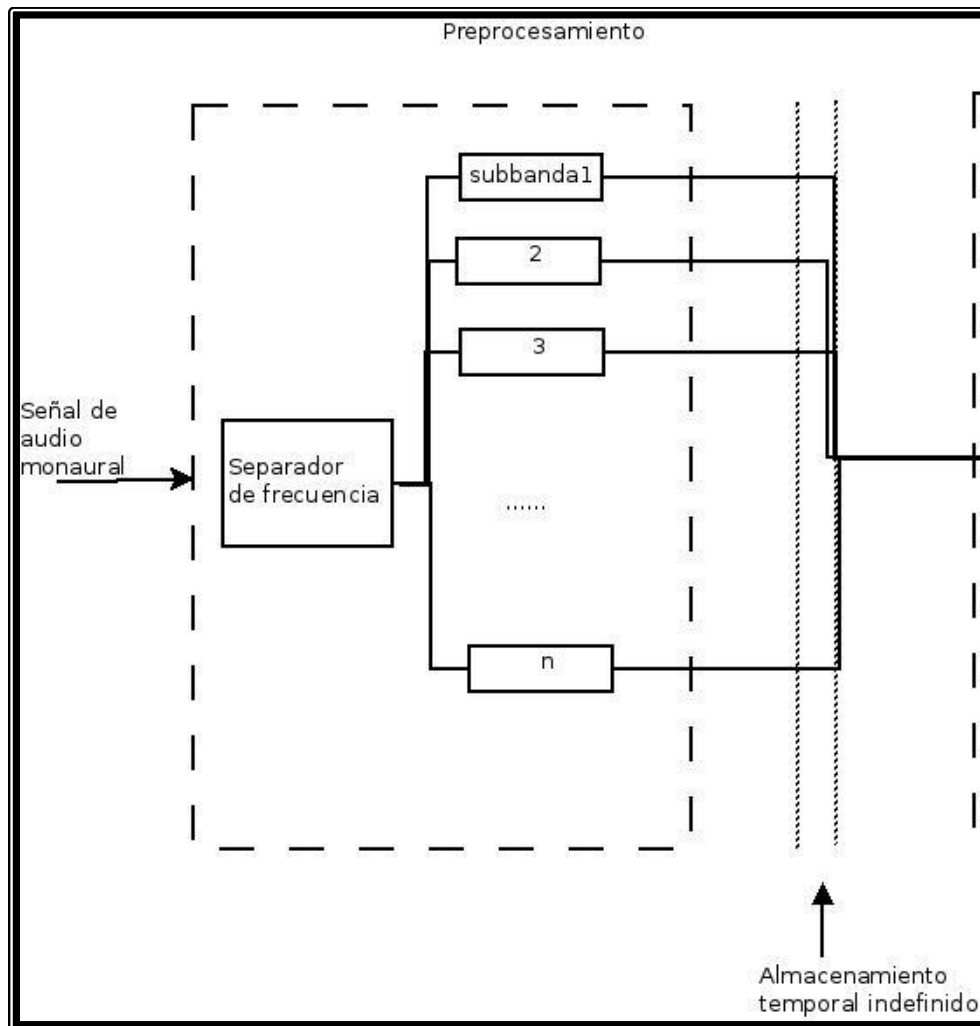


Figura 6.1 Predescomposición en n bandas de frecuencia de la señal de audio de entrada. La salida puede almacenarse en un medio de manera indefinida para su uso posterior o bien introducirse directamente a la segunda etapa del sistema (etapa de generación en tiempo real).



ETAPA DE PROCESAMIENTO EN TIEMPO REAL

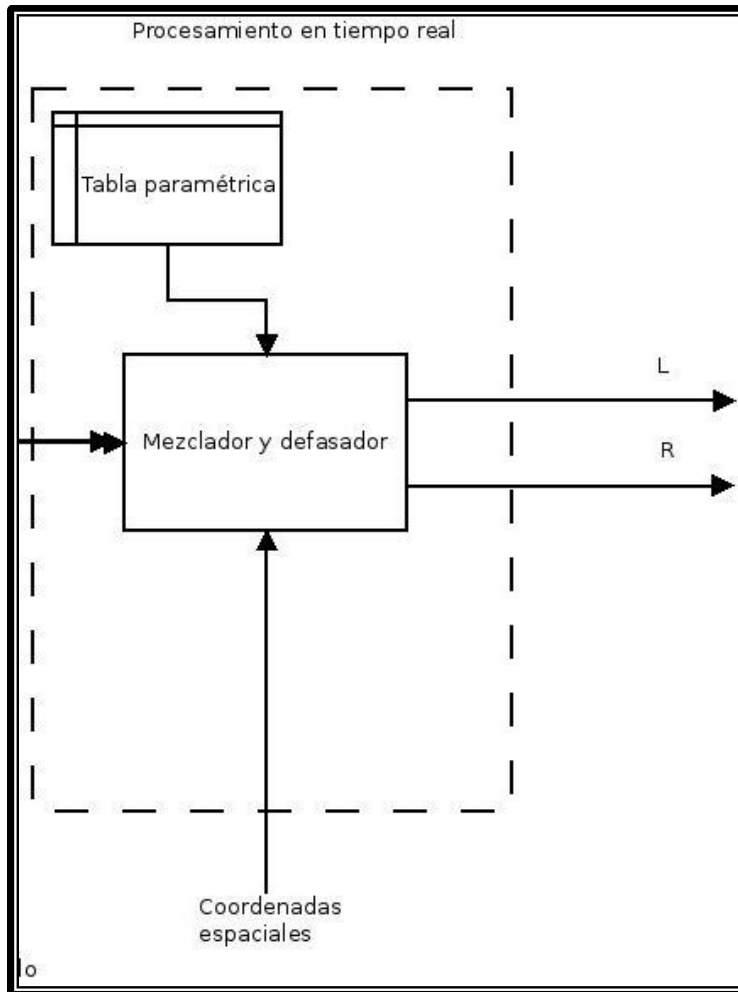


Figura 6.2 Etapa de procesamiento en tiempo real. La entrada a esta etapa son las coordenadas espaciales, las n subbandas de frecuencia de la etapa de predescomposición así como los datos obtenidos experimentalmente de la tabla paramétrica.



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



6.3 Implementación final

Dado el diagrama de bloques, a continuación colocamos las distintas herramientas que se desarrollaron como parte de este trabajo.

Filtro paso bajas basado en un simple promedio

Para fines demostrativos de la técnica se usó un simple filtro que promedia el valor de la señal de audio cada λ (período completo o una longitud de onda completa), de tal forma que se obtienen frecuencias por debajo de la frecuencia de corte del filtro paso bajas. La implementación de un filtro paso bandas se hará a partir de este filtro.

```
#include <defs.h>
int main(int argc, const char *argv[]){
    int *outsamples;
    int *insamples;
    int i, j, k, n, avg, avgv, avgvp;
    char skip = FALSE;
    SNDFILE *inputf;
    SF_INFO input;
    SNDFILE *outputf;
    SF_INFO output;

    if(argc<4){
        printf("Usage: %s input.wav output.wav
samples_to_average\n",argv[0]);
```



CAPÍTULO 6



```
        return 0;
    }
    printf("Abriendo archivo %s\n", argv[1]);
    input.format = 0;
    inputf = sf_open(argv[1], SFM_READ, &input);
    if (inputf == NULL)
    {
        printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[1]);
        goto exit;
    }
    n = input.frames * input.channels;
    insamples = (int *)malloc(sizeof(int)*n);

    if (sf_read_int(inputf, insamples, n) != n)
    {
        printf("No se pudo leer las muestras del archivo
de entrada %s\n", argv[1]);
        goto exit;
    }

    output.samplerate = input.samplerate;
    output.channels = input.channels;
    output.format = SF_FORMAT_WAV | SF_FORMAT_PCM_16;
    printf("Abriendo archivo %s\n", argv[2]);
    outputf = sf_open(argv[2], SFM_WRITE, &output);
    if (outputf == NULL)
    {
        printf("Error: No se pudo abrir para escritura
el archivo %s\n", argv[2]);
        goto exit;
    }
}
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
    }
    outsamples = (int *)malloc(sizeof(int)*n);
    memset(outsamples, 0, sizeof(int)*n);
    avg = atoi(argv[3]); //Valor entero
    //Sacar el promedio y ponerlo en la señal de
    salida
    for(i = 0; i < n; i += avg){
        avgv = 0;
        for(j=i; ( j < i + avg ) && ( j < n );j++)
        {
            avgv += insamples[j]/avg;
        }
        //Listo, ponerlo en la salida
        for( j = ( i - avg/2); ( j < i + avg/2 + 2 ) && (
j < n ) ;j++)
        {
            if( i < (avg/2 + 2))
                outsamples[j] = avgv;
            else
            {
                // Interpolación lineal
                //outsamples[j] = avgvp + (int)((float)j-
i+avg/2)*(avgv - avgvp)/avg;
                // Interpolación senoidal
                outsamples[j] = avgvp + (int)((cos((j-
i+avg/2)*PI/avg))*(avgvp - avgv)/2) + (avgv-avgvp) /
2;

            }
        }
        avgvp = avgv;
    }
```



CAPÍTULO 6



```
    }

    output.frames = n;
    sf_write_int(outputf, outsamples, n);

exit:

    if(inputf) sf_close(inputf);
    if(outputf) sf_close(outputf);
    if(outsamples) free(outsamples);
    if(insamples) free(insamples);

}
```

Para utilizarlo en la línea de comandos basta con hacer lo siguiente:

```
./averager entrada.wav salida.wav muestras_a_promediar
```

Por ejemplo, para filtrar alrededor de 100 Hz, el número de muestras en una señal de 44.1kHz que comprenden un período completo es

$$\lambda = \frac{f_s}{f_c} = \frac{44100}{100} = 441$$



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



Por tanto,

```
./averager input.wav output110.wav 441
```

Producirá una onda de sonido reducida a 100 Hz como frecuencia de corte. En el código puede notarse que, debido a que la señal de salida también es de 44.1 kHz, es necesario realizar interpolación para las muestras intermedias. En la experimentación, una interpolación lineal demostró introducir ruido molesto de alta frecuencia, por lo que se optó por una interpolación trigonométrica cosenoidal de primer orden (un solo polinomio trigonométrico). Esto garantiza que habrá poco contenido espectral encima de la frecuencia de corte, con la desventaja de introducir una componente cosenoidal en la frecuencia de corte, no existente en la señal original. En realidad esto no es muy importante, ya que se pueden usar filtros posteriores de alto factor de calidad (supresor de banda) para eliminar dicha espiga o bien, usar en una implementación otro filtro.

Restador de señales de audio

La idea detrás de esta herramienta es ser el auxiliar para crear, junto con el filtro paso bajas, un filtro paso bandas sencillo que sirva para ejemplificar la técnica desarrollada.



CAPÍTULO 6



```
#include <defs.h>

int main(int argc, const char *argv[]){
    int *outsamples;
    int *insamples;
    int *insamples2;
    int i, n, n2, no;
    SNDFILE *inputf;
    SNDFILE *inputf2;
    SF_INFO input;
    SF_INFO input2;
    SNDFILE *outputf;
    SF_INFO output;

    if(argc<4){
        printf("Usage: %s minuendo.wav sustraendo.wav
resta.wav\n", argv[0]);
        return 0;
    }
    printf("Abriendo archivo %s\n", argv[1]);
    input.format = 0;
    inputf = sf_open(argv[1], SFM_READ, &input);
    if (inputf == NULL)
    {
        printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[1]);
        goto exit;
    }
    printf("Abriendo archivo %s\n", argv[2]);
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
input2.format = 0;
inputf2 = sf_open(argv[2], SFM_READ, &input2);
if (inputf2 == NULL)
{
    printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[1]);
    goto exit;
}
n = input.frames * input.channels;
n2 = input2.frames * input2.channels;
no = (input.frames * input.channels > input2.frames
* input2.channels) ? input.frames *
input.channels:input2.frames * input2.channels;
insamples = (int *)malloc(sizeof(int)*n);
insamples2 = (int *)malloc(sizeof(int)*n2);

if (sf_read_int(inputf, insamples, n) != n)
{
    printf("No se pudo leer las muestras del archivo
de entrada %s\n", argv[1]);
    goto exit;
}
if (sf_read_int(inputf2, insamples2, n2) != n2)
{
    printf("No se pudo leer las muestras del archivo
de entrada %s\n", argv[2]);
    goto exit;
}

output.samplerate = input.samplerate;
output.channels = input.channels;
```



CAPÍTULO 6



```
output.format = SF_FORMAT_WAV | SF_FORMAT_PCM_16;
printf("Abriendo archivo %s\n",argv[3]);
outputf = sf_open(argv[3], SFM_WRITE, &output);
if (outputf == NULL)
{
    printf("Error: No se pudo abrir para escritura el
archivo %s\n", argv[3]);
    goto exit;
}
outsamples = (int *)malloc(sizeof(int)*no);
memset(outsamples, 0, sizeof(int)*no);
for(i = 0; i < no; i++){
    outsamples[i] = insamples[i] - insamples2[i];
}

output.frames = no;
sf_write_int(outputf, outsamples, no);

exit:

if(inputf) sf_close(inputf);
if(outputf) sf_close(outputf);
if(outsamples) free(outsamples);
if(insamples) free(insamples);

}
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



Por ejemplo, para obtener una señal de entre 400 y 1200 Hz, se filtra primero la señal original en 400 y 1200 Hz usando el filtro paso bajas, y luego se resta la señal filtrada hasta 400 Hz a la señal de 1200 Hz:

```
./averager input.wav output400.wav 110
./averager input.wav output1200.wav 37
./restador output1200.wav outout400.wav 400to1200.wav
```

Las dos herramientas anteriormente generadas corresponden a la etapa de predescomposición en bandas de frecuencias. Las siguientes implementan en su conjunto la etapa de procesamiento en tiempo real.

Amplificador

Se escribió esta herramienta para ser capaces de alterar la ganancia de un flujo de sonido dado. A continuación se incluye el código fuente para compilar dicha herramienta:

```
#include <defs.h>

int main(int argc, const char *argv[]){
    int *outsamples;
    int *insamples;
    int i;
    int n;
    double gain;
```



CAPÍTULO 6



```
SNDFILE *inputf;
SF_INFO input;
SNDFILE *outputf;
SF_INFO output;

if(argc<4){
    printf("Usage: %s input.wav output.wav
gain_in_db\n",argv[0]);
    return 0;
}
printf("Abriendo archivo %s\n", argv[1]);
input.format = 0;
inputf = sf_open(argv[1], SFM_READ, &input);
if (inputf == NULL)
{
    printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[1]);
    goto exit;
}
n = input.frames * input.channels;
insamples = (int *)malloc(sizeof(int)*n);

if (sf_read_int(inputf, insamples, n) != n)
{
    printf("No se pudo leer las muestras del archivo
de entrada %s\n", argv[1]);
    goto exit;
}

output.samplerate = input.samplerate;
output.channels = input.channels;
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
output.format = SF_FORMAT_WAV | SF_FORMAT_PCM_16;
printf("Abriendo archivo %s\n",argv[2]);
outputf = sf_open(argv[2], SFM_WRITE, &output);
if (outputf == NULL)
{
    printf("Error: No se pudo abrir para escritura
el archivo %s\n", argv[2]);
    goto exit;
}
gain = pow(10,atoi(argv[3])/(double)10);

printf("Gain used: %e\n", gain);
outsamples = (int *)malloc(sizeof(int)*n);

for(i=0;i<n;i++)
{
    outsamples[i] = gain*insamples[i];
}
output.frames = n;
sf_write_int(outputf, outsamples, n);

exit:

if(inputf) sf_close(inputf);
if(outputf) sf_close(outputf);
if(outsamples) free(outsamples);
if(insamples) free(insamples);

}
```



CAPÍTULO 6



La manera de utilizarlo en la línea de comandos es como sigue:

```
./amp input.wav output.wav gain_in_dB
```

Por ejemplo:

```
./amp 100Hz.wav 100Hz3dB.wav 3
```

Generará el archivo 100Hz3dB.wav a partir del archivo 100Hz.wav, dándole una ganancia de 3 dB.

Generador de retraso

Herramienta utilizada para generar el retraso deseado. Dada una señal de audio basta con introducir el archivo de salida y el número de muestras a retrasar con respecto a la señal original. Si se retrasa, por ejemplo, un segundo, se llena con ceros el segundo antes de iniciar la reproducción y asimismo se agranda la duración del archivo para que no haya corte al final.

```
#include <defs.h>

int main(int argc, const char *argv[]){
    int *outsamples = NULL;
```




IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
int *insamples = NULL;
int i;
int n;
int t;
int delay;
SNDFILE *inputf = NULL;
SF_INFO input;
SNDFILE *outputf = NULL;
SF_INFO output;

if(argc<4){
    printf("Usage: %s input.wav output.wav
delay_in_frames\n",argv[0]);
    return 0;
}
printf("Abriendo archivo %s\n", argv[1]);
input.format = 0;
inputf = sf_open(argv[1], SFM_READ, &input);
if (inputf == NULL)
{
    printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[1]);
    goto exit;
}
n = input.frames * input.channels;
insamples = (int *)malloc(sizeof(int)*n);

if ((t = sf_read_int(inputf, insamples, n)) != n)
{
    printf("No se pudo leer las muestras del archivo
de entrada %s\n", argv[1]);
```



CAPÍTULO 6



```
        goto exit;
    }
    //Hardcoding
    output.samplerate = input.samplerate;
    output.channels = input.channels;
    output.format = SF_FORMAT_WAV | SF_FORMAT_PCM_16;
    printf("Abriendo archivo %s\n",argv[2]);
    outputf = sf_open(argv[2], SFM_WRITE, &output);
    if (outputf == NULL)
    {
        printf("Error: No se pudo abrir para escritura el
archivo %s\n", argv[2]);
        goto exit;
    }
    delay = atoi(argv[3])*output.channels;
    n = (input.frames*input.channels) + delay;
    outsamples = (int *)malloc(sizeof(int)*n);
    memset(outsamples, 0, sizeof(int)*delay);
    //Create the delay

    n = input.frames * input.channels;
    for(i=0;i<n;i++)
    {
        outsamples[i+delay] = insamples[i];
    }
    output.frames = input.frames + atoi(argv[3]);
    n = (input.frames*input.channels) + delay;
    sf_write_int(outputf, outsamples, n);

exit:
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
if(inputf) sf_close(inputf);  
if(outputf) sf_close(outputf);  
if(outsamples) free(outsamples);  
if(insamples) free(insamples);  
  
}
```

Su uso es similar al de las demás herramientas:

```
./delay input.wav output.wav delay_in_frames
```

Izquierda y derecha

Dos herramientas que no vale mucho la pena incluir pero que se desarrollaron, generan una señal de audio estéreo, copia de una señal monaural, y coloca la señal monaural en uno de los dos canales. Esto es importante para la simulación ya que se necesita generar una copia de la señal original para cada subbanda de frecuencia y aplicarle efectos distintos, según se trate del canal izquierdo o del derecho.

Mezclador

Una vez alteradas las n subbandas de frecuencia para los canales izquierdo y derecho, sólo queda mezclarlos para producir la salida final del sistema. El siguiente código lleva a cabo dicha operación.



CAPÍTULO 6



```
#include <defs.h>

int main(int argc, const char *argv[]){
    char filename[100];
    int n;
    int *outsamples;
    int *insamples[5];
    int i;
    SNDFILE *inputsf[5];
    SF_INFO inputs[5];
    SNDFILE *outputf;
    SF_INFO output;
    double precomputed;

    if(argc<3){
        printf("Usage: %s file1.wav file2.wav [file3.wav
... file_n.wav]\n",argv[0]);
        printf("The output will be a mix of all those
audio signals. The name of the output will be
asked.\n");
        printf("For simple reasons, max inputs are 5. If
you need more mixes, mix 5 by 5\n");
        return 0;
    }
    //Ask for the output name
    printf("Introduce el nombre del archivo de salida:
");
    scanf("%s", filename);
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
//Get Input files and read info
for(i = 0; i< argc - 1; i++){
    printf("Abriendo archivo %s\n",argv[i+1]);
    inputsf[i] = sf_open(argv[i+1], SFM_READ,
&inputs[i]);
    if (inputsf[i] == NULL)
    {
        printf("Error: No se pudo abrir para lectura el
archivo %s\n", argv[i+1]);
        goto exit;
    }
    insamples[i] = (int
*)malloc(sizeof(int)*inputs[i].frames*inputs[i].chann
els);

    if (sf_read_int(inputsf[i], insamples[i],
inputs[i].frames*inputs[i].channels) !=
inputs[i].frames*inputs[i].channels)
    {
        printf("No se pudo leer las muestras del
archivo de entrada %s\n", argv[i+1]);
        printf("input frames: %d\n", (int)
inputs[i].frames);
        goto exit;
    }
}
//Hardcoding
output.samplerate = inputs[0].samplerate;
output.channels = inputs[0].channels;;
output.format = SF_FORMAT_WAV | SF_FORMAT_PCM_16;
```



CAPÍTULO 6



```
outputf = sf_open(filename, SFM_WRITE, &output);
if (outputf == NULL)
{
    printf("Error: No se pudo abrir para escritura el
archivo %s\n", filename);
    goto exit;
}
//Determine the max number of frames.
output.frames=inputs[0].frames;
for(i=0;i<argc-1;i++)
{
    if(inputs[i].frames > output.frames)
output.frames = inputs[i].frames;
}
outsamples = (int
*)malloc(sizeof(int)*output.frames*output.channels);
memset(outsamples, 0,
sizeof(int)*output.frames*output.channels);
//Mix everything

for(i=0;i<output.frames*output.channels;i++)
{
    outsamples[i] = 0;
    for(n=0;n<argc-1;n++){
//        if (inputs[n].frames*inputs[n].channels > i )
//        I think we should not divide by the number of
signals
//        outsamples[i] += insamples[n][i]/(argc-1);;
        outsamples[i] += insamples[n][i];
    }
}
}
```



IMPLEMENTACIÓN DEL MODELO EN UN SIMULADOR DE TIEMPO REAL



```
sf_write_int(outputf, outsamples, i);

exit:

for(n=0;n<argc-1;n++){
    if (inputsf[n])
        sf_close(inputsf[n]);
}
if(outputf) sf_close(outputf);
if(outsamples) free(outsamples);
for(n=0;n<argc-1;n++){
    if (insamples[n])
        free(insamples[n]);
}

}
```

Su uso es bastante sencillo:

```
./mixer s1.wav s2.wav s3.wav.....sn.wav output.wav
```

Con esto se consigue la señal de audio con los efectos necesarios para percibir el sonido desde una fuente proveniente de cualquier lugar, siempre y cuando las herramientas de ganancia y retraso en el tiempo se utilicen para cada canal y con los parámetros adecuados.



CAPÍTULO 6



CONCLUSIONES



7.1 Comentarios de la implementación

Si bien la implementación no utiliza un dispositivo de E/S en tiempo real, los tiempos de ejecución de todas las herramientas utilizadas nunca superaron el tiempo de reproducción de los archivos de sonido utilizados.

Lo que se ha hecho en el trabajo presente es que la carga de trabajo para generar el efecto de audio envolvente se ha dividido en dos etapas. Una es una etapa de preprocesamiento, la cual busca generar n señales, subbandas a partir de la señal original de entrada. Esta etapa, aunque no es tan intensa computacionalmente hablando, requiere de que exista al menos una cantidad de muestras en la señal suficientes para obtener un período completo de una onda de frecuencia igual a 20 Hertz, es decir de 50 ms, de la señal M para producir las subbands. Esto significa que la solución no contempla ir en tiempo real de la señal original a las señales de salida, sino que solamente se puede ir en tiempo real desde las subbandas $B1$ a las señales a reproducir en cada oído.

El problema de generar audio envolvente a partir de una señal cualquiera se ha resuelto aquí tomando en cuenta las siguientes consideraciones:

1.- El uso aumentado de memoria en el sistema destino no es limitante. Esto permite preprocesar la señal original en n señales y almacenarlas como parte de la aplicación.

2.- La generación en tiempo real, con la posición como una entrada en tiempo real de la función generadora, debe ser posible en situaciones donde el recurso computacional de velocidad de procesamiento es limitado. Aquí se ha logrado



CONCLUSIONES



minimizar esta función a $2n$ operaciones de entradas reales con n elementos de retraso para la señal izquierda y una cantidad similar y susceptible de optimizarse compartiendo recursos para generar la señal derecha. Una implementación en particular puede optimizar casos particulares y algunos detalles. Si bien se puede ver que la complejidad computacional de la etapa de mezclado final es lineal con respecto a la cantidad de señales n de la etapa de descomposición en subbandas, el número de operaciones es fijo con respecto a cualquier parámetro de la señal original. Si, por ejemplo, la salida de la etapa de descomposición son 8 señales n , subbandas de la señal original, el número total de operaciones para generar las señales izquierda y derecha, es 48, número fijo e independiente de la señal de entrada. (8 multiplicaciones para la ganancia, 8 accesos de memoria para los elementos de retraso y 8 sumas para la mezcla final, multiplicado todo esto por 2, para las dos señales de salida, da las 48 operaciones aritméticas de números reales). En un procesador comercial móvil, para un smartphone de 1 GHz, generar una muestra de dichas señales tomaría alrededor de 48 nanosegundos. No es este el caso con soluciones que utilizan conversión al dominio de la frecuencia, cuya complejidad computacional es mayor.

3.- La información de la posición de la fuente tiene limitaciones sobre su velocidad de cambio. Cambiar la posición de la fuente de formas que superen la realidad (como un cambio instantáneo de más de un metro o velocidades por encima de lo cotidianamente esperado) puede llegar a generar “artefactos” desagradables, principalmente clipping (chasquidos en la señal). Además de limitar la brusquedad de cambio en la posición de la fuente de sonido (asunto de una implementación particular), se debe considerar, debido a esto, un procesamiento adicional de suavizado, si es que puede suceder algún artefacto de sonido.



7.2 Restricciones de la implementación

La implementación depende de la preexistencia de la señal de entrada. Esto significa que toda la aplicación no es en tiempo real. Se puede generar audio envolvente interactivo en tiempo real respecto a la posición que se pretende simular. Es decir, la posición a simular puede generarse arbitrariamente en tiempo real.

Otra restricción importante es que la etapa en tiempo real requiere que la señal este en su forma de onda, ya que por la naturaleza del proceso no se puede trabajar con una señal de audio que esté en el dominio de la frecuencia. Esto, sin embargo, no es una limitante demasiado restrictiva, ya que este efecto puede implementarse al final de todo el procesamiento de audio, cuando la señal de salida está lista para reproducirse y se encuentra representada en su forma de onda.

7.3 Ventajas y desventajas

La principal ventaja de esta técnica radica en que es fácil de implementar además de tener un muy bajo costo en recursos de cómputo. De la misma manera, el diseño del sistema es altamente paralelizable, tanto para subdividir el problema de generar una sola señal en distintas unidades de trabajo como para generar varias señales con distintas ubicaciones virtuales, aprovechando la linealidad del sistema.



CONCLUSIONES



Sin embargo, el procesamiento general a la señal de audio de entrada puede llegar a introducir efectos o ruidos no deseados, razón por la cual también es necesario ajustar la implementación para la aplicación objetivo. Esto incluye agregar filtros intermedios para suavizar la señal de salida en términos de ruidos generados por el sistema.

Otra desventaja de esta metodología radica en que necesariamente requiere un sistema de almacenamiento para las n subbandas. Sin dicho almacenamiento, tendría que considerarse que la etapa de preprocesamiento es también en tiempo real, y esto podría significar una mayor complejidad en el diseño general del sistema.

REFERENCIAS BIBLIOGRÁFICAS



REFERENCIAS BIBLIOGRÁFICAS



Título: Genetics, a conceptual approach

Autor: Benjamin A. Pierce

Editorial: W.H. Freeman

Título: Signals and Systems

Autor: Alan V. Oppenheim

Editorial: Prentice Hall

Título: The Human Brain: An introduction to its Functional Anatomy

Autor: John Nolte

Editorial: Mosby Elsevier

Título: Master Handbook of Acoustics

Autor: F. Althon Everest

Editorial: McGraw Hill

Título: Mixing Audio. Concepts, Practices and Tools

Autor: Roey Izhaki

Editorial: Focal Press

Título: Introduction to Digital Filters: with Audio Applications

Autor: Julius O. Smith III

Editorial: Booksurge publishing

Título: C Programming Language

Autor: Brian W. Kernighan/ Dennis M. Ritchie

Editorial: Prentice Hall Software Series