

UNIVERSIDAD POLITECNICA DE VALENCIA

ESCUELA POLITECNICA SUPERIOR DE GANDIA

Master en Ingeniería Acústica

---



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



ESCUELA POLITECNICA  
SUPERIOR DE GANDIA

**“Procesador de sonido y estudio de  
métodos para la generación de audio  
3D en tiempo real”**

*TESIS DE MASTER*

Autor:

*Pablo Benedito Botella*

Director/es:

*D. Jorge Gosálbez Castillo*

*GANDIA, 2012*

# Procesador de sonido y estudio de métodos para la generación de audio 3D en tiempo real

*Autor:* Pablo Benedito Botella

*Director:* Jorge Gosálbez Castillo

*Resumen — La reproducción del sonido en 3D está incrementando su importancia dentro del mundo científico, comercial y del ocio, por lo que la demanda de sistemas capaces de operar en tiempo real aumenta cada vez más. Para agilizar los procesos de escucha 3D, es necesario que se implementen funciones que permitan la generación y adaptación de estos procesos en tiempo real. Gracias a este procesado en tiempo real, la adaptación se puede hacer en función de la posición del emisor y receptor junto con las condiciones de la sala creando un espacio acústico virtual en el que tanto el emisor como el receptor puedan realizar desplazamientos. En este trabajo se estudian diferentes teorías y métodos para la generación de sonido 3D (auralización) así como la realización de una campaña de medidas para evaluar los resultados de estos efectos. Como novedad se presenta la adaptación de estos efectos al movimiento del oyente en dos etapas: una primera etapa con un movimiento simulado y una segunda mediante un posicionador o Tracker. Esto permite que los sonidos 3D se adapten a los movimientos del oyente y que, por tanto, mantenga la sensación de inmersión en el espacio acústico virtual a pesar que realice algún tipo de desplazamiento.*

*Abstract — The 3D sound reproduction is increasing its importance in the scientific, commercial and leisure world, so that the demand for systems capable of operating in real time is constantly increasing. To streamline 3D listening processes, it is necessary to implement functions that allow the generation and adaptation of these processes in real time. Through this real-time processing, adaptation can be realized depending on the position of the transmitter and receiver along with the conditions of the room creating a virtual acoustic space in which both the sender and the receiver can perform movements. In this thesis we study different theories and methods for the generation of 3D sound (auralization) and conducting a campaign of measures to assess the results of these effects. As a novelty, the adaptations to these effects is presented to the movement of the listener in two stages: a first stage with a simulated motion and a second by a positioner or Tracker. This allows 3D sounds to adapt to the listener's movements and therefore maintain the feeling of immersion in the virtual acoustic space despite performing some type of displacement.*

## ÍNDICE

<b>I. Introducción</b> .....	<b>3</b>
I.1. Objetivos.....	3
I.2. Conceptos básicos de la escucha del sonido 3D.....	4
<b>II. Teoría de la holofonía del sonido</b> .....	<b>5</b>
II.1. Introducción.....	5
II.2. Sistema de referencia.....	7
II.3. Determinación del ángulo lateral (azimut).....	7
II.4. Determinación de la elevación .....	9
II.5. Ambigüedad en la localización.....	10
II.6. Head-related transfer functions (HRTF).....	11
<b>III. Externalización del sonido</b> .....	<b>13</b>
III.1. Teoría para la externalización del sonido .....	13
III.2. Modelización con la distancia.....	14
<b>IV. Selección del modelo y lenguaje de desarrollo</b> .....	<b>16</b>
IV.1. Elección del lenguaje de desarrollo .....	16
IV.2. Implementación de la holofonía: HRTF.....	17
IV.3. Implementación de la externalización (reverberación): RIR.....	19
IV.4. Integración de las HRTF y RIR.....	23
<b>V. Generación del efecto de auralización en tiempo real</b> .....	<b>24</b>
V.1. Introducción.....	24
V.2. Elección del tracker .....	25
V.3. Desarrollo de la aplicación .....	29
V.4. Aplicaciones .....	32
<b>VI. Campaña de medidas</b> .....	<b>32</b>
VI.1. Introducción.....	32
VI.2. Test HRTF .....	34
VI.3. Test RIR o reverberación.....	36
VI.4. Test final .....	37
<b>VII. Conclusiones y líneas futuras</b> .....	<b>39</b>
<b>AGRADECIMIENTOS</b> .....	<b>40</b>
<b>BIBLIOGRAFÍA</b> .....	<b>41</b>

## I. INTRODUCCIÓN

### I.1. OBJETIVOS

La auralización de la fuente monofónica de sonido, nos permite ubicar virtualmente la fuente de sonido en una orientación, distancia y tipo de sala. De esta manera, el oyente cuando escucha la grabación auralizada mediante unos auriculares, tiene la sensación acústica que dicho sonido le viene desde un punto determinado de una determinada sala. Dicha auralización la podemos considerar estática o dinámica. En el primer caso, tanto oyente como fuente están estáticos uno respecto a otro, mientras que en el segundo, hay un desplazamiento relativo entre ambos, normalmente asociado a la fuente. En este trabajo pretendemos avanzar en estas tecnologías de realidad aumentada, relacionadas con el audio.

Para ello estudiaremos e implementaremos diferentes algoritmos para la auralización de fuentes monofónicas, donde, cabe destacar la aplicación en tiempo real y en función de la posición del oyente. Este último aspecto supone un avance ya que permitirá independizar el movimiento del oyente respecto a la fuente de sonido. Como contrapartida requerirá de un sistema de posicionamiento y orientación de la persona dentro del recinto o sala virtual.

Para alcanzar la auralización estudiaremos en primer lugar las técnicas para posicionar virtualmente la fuente de audio respecto a la cabeza del oyente. Para este objetivo, nos centraremos en algoritmos basados en las HRTF (*Head-Related Transfer Function*) (“2. Teoría de la homofonía del sonido”), las cuales modelan la respuesta al impulso de cada oído en dependiendo de la orientación entre la fuente acústica y la cabeza del oyente.

En segundo lugar, necesitaremos también modelar la distancia entre la fuente de audio y el oyente con el fin de externalizar dicha fuente. Para ello, en el capítulo “3. Externalización del sonido” hacemos una búsqueda bibliográfica de los algoritmos de reverberación de salas los cuales modelan la respuesta al impulso de la sala entre un punto transmisor y otro receptor. Tras dicho, estudio centraremos nuestro trabajo en los algoritmos de reverberación basados en el posicionamiento de fuentes imagen.

Finalmente, la combinación de estos algoritmos junto con integración de la posición del oyente en tiempo real, permitirá adaptar la posición de la fuente acústica en función del tipo de sala y de la orientación del oyente. Para lograrlo tendremos que buscar una tecnología adecuada (*tracker* o posicionador, lenguaje de programación, interfaz de desarrollo,...) (“4. Selección del modelo y lenguaje de desarrollo”) que permita desarrollar nuestro objetivo, es decir, la

aplicación de la auralización adaptada a la posición del oyente en tiempo real (“5. Generación del efecto de auralización en tiempo real”).

Todo esto será validado y evaluado en la última fase del proyecto mediante una campaña de test (“6. Campaña de medidas”).

Finalmente, en el capítulo, “7. Conclusiones y líneas futuras”, expondremos las principales conclusiones de este trabajo, las futuras líneas abiertas así como diferentes campos de aplicación para esta tecnología desarrollada.

## *1.2. CONCEPTOS BÁSICOS DE LA ESCUCHA DEL SONIDO 3D*

Uno de los elementos clave que separa el procesado de señales de audio del procesado general de señales es que, en las señales de audio, el receptor final es siempre un ser humano. Por lo tanto, todas las operaciones hechas a la señal, no sólo deben de ser investigadas por los medios tradicionales, sino también desde el punto de vista de las capacidades perceptivas del sistema auditivo humano.

La audición es un proceso complejo. El cerebro humano, para interpretar un sonido, ha de procesar la información que le llega de ambos oídos. La información que el cerebro recibe de cada uno de los oídos es diferente, porque ambos oídos están físicamente separados entre sí por la cabeza. Mediante esta diferencia en la situación de los oídos, el cerebro es capaz de situar en el espacio cualquier fuente sonora. Para ello es necesaria la audición binaural, ya que con un solo oído no sería posible esta localización. Por lo tanto la fisiología de cada persona, así como las propiedades de la fuente y las características de la sala en la que se realiza el proceso de escucha son varios de los factores de los que depende el correcto posicionamiento espacial.

Cada configuración anatómica de cada oyente realiza un filtrado y unas reflexiones propias. La función de transferencia que relaciona el sonido que llega al tímpano con el emitido por la fuente sonora, se denomina HRTF (*Head-Related Transfer Function*). Ésta también puede ser sintetizada, desarrollando un modelo estructural relativamente simple cuyos parámetros están estrictamente relacionados con las características fisiológicas de cada individuo.

La correcta simulación del HRTF permite crear la ilusión de espacialidad en el cerebro, permitiendo que se pueda localizar la dirección de llegada de una onda sonora determinada.

Cuando un sonido proviene de algún lugar del espacio, una persona es capaz de localizar su procedencia en azimut, elevación y distancia. Sin duda, la estimación de su posición en azimut es mucho más precisa que el resto, y la de su posición en la distancia es la más difícil de todas.

El otro parámetro necesario para la localización espacial del sonido es la distancia. La señal que vamos a percibir depende directamente de las características de la sala donde nos encontremos, ya que ésta se va a modificar según el tipo de recinto en el que nos encontremos. Cada sala tiene asociada una función de transferencia propia, denominada respuesta al impulso ( $h_{sala}(t, x_{emisor}, x_{receptor})$ ). Este parámetro modela las características de la sala en función del punto del emisor y el punto del receptor. Por eso si variamos tanto la posición de la fuente o la del receptor, la función de transferencia va a ser diferente.

La correcta simulación de la reverberación de la sala permite crear la ilusión de distancia, permitiendo que se pueda determinar la separación respecto del oyente de una fuente sonora determinada.

## II. TEORÍA DE LA HOLOFONÍA DEL SONIDO

### II.1. INTRODUCCIÓN

Las diferentes técnicas empleadas para la recreación tridimensional de una fuente sonora se denominan holofonía. Esto es, al igual que en un holograma, conseguir recrear la sensación en el oyente de un sonido proveniente de una determinada dirección, y a una determinada distancia en el espacio.

Hay varios métodos de realizar estas grabaciones holofónicas, que registran información direccional: mediante el uso de un maniquí acústico o modelo artificial, a través de un voluntario que lleve un par de micrófonos especialmente diseñados para colocarse en el interior del oído, para tener así un auténtico modelo real del efecto del cuerpo sobre el sonido, o también se puede simular electrónicamente el efecto de la cabeza para incorporar la información sobre la dirección al sonido.

Las primeras grabaciones que consiguen este efecto fueron realizadas por el investigador ítalo-argentino Hugo Zucarelli, quien para ello utilizó un maniquí acústico o *dummy* que reproducía fielmente una cabeza y torso humano con las mismas características acústicas que el cuerpo humano. Así, con dos micrófonos situados en los oídos del maniquí, se obtenía una grabación que, al escucharla con unos auriculares, conseguía el efecto deseado. El principal

inconveniente de estas grabaciones es la falta de interactividad, ya que la posición del *dummy* en el momento de la grabación respecto a la fuente sonora era la determinante de la sensación que percibiría el oyente [1].

Mediante el uso de maniqués se han realizado grabaciones que consiguen efectos enormemente reales, y se han empleado incluso de forma comercial en grabaciones de música y sonidos relajantes. Se recomienda cerrar los ojos para así evitar la confusión provocada al no ver una fuente del sonido real, ya que nuestro cerebro se apoya en la información visual para identificar la fuente sonora.

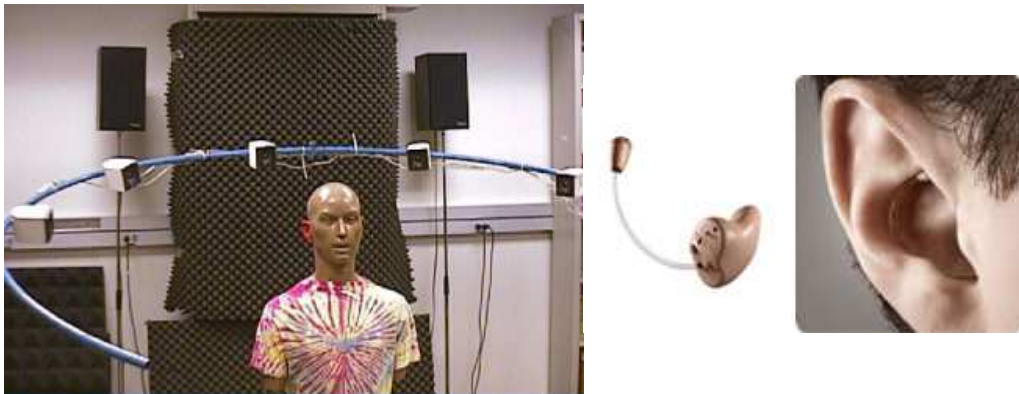


Figura 1. Métodos de medida de HRTF mediante maniqués acústicos o micrófonos en el oído.

Otro método es poner unos micrófonos especiales a una persona para realizar una grabación real. Éste método tiene la principal ventaja de la movilidad del voluntario, pudiéndose realizar grabaciones que simulen situaciones más reales y que incluyen movimiento. Por otro lado, este sistema tiene también desventajas, ya la grabación está más sujeta a las características del voluntario, limitando la compatibilidad entre el voluntario que realiza la grabación y quien finalmente la escucha, debido a las diferencias anatómicas de sus cabezas y pabellones auditivos, perjudicando la fidelidad de la grabación cuando otra persona lo escuche.

Sin embargo, con el desarrollo de la computación han aparecido sistemas informáticos capaces de reproducir el efecto de la cabeza y torso humanos sobre el sonido que percibimos, recreando el efecto artificialmente. Son los llamados sistemas basados en HRTF (*Head-Related Transfer Function*). Gran parte de la investigación actual se centra en el análisis y determinación de estas funciones, y de su caracterización mediante un conjunto limitado de parámetros [2].

## II.2. SISTEMA DE REFERENCIA

Para poder determinar la ubicación de una fuente sonora en el espacio, el sistema auditivo utiliza un conjunto de pistas que es necesario situar dentro de un sistema de coordenadas para poder estudiarlas. Como la cabeza puede aproximarse con una esfera, se suele utilizar como origen el centro de la cabeza y el acimut, la elevación y la distancia como variables que definen cualquier punto del espacio alrededor de esta. La forma más usual de definir dichas coordenadas es el sistema de coordenadas vertical-polar. Con esta elección, las superficies de acimut constante son planos a través del eje Z y las superficies de elevación constante son conos concéntricos alrededor del eje Z [3].

A continuación se muestra una figura que aclara un poco más el concepto:

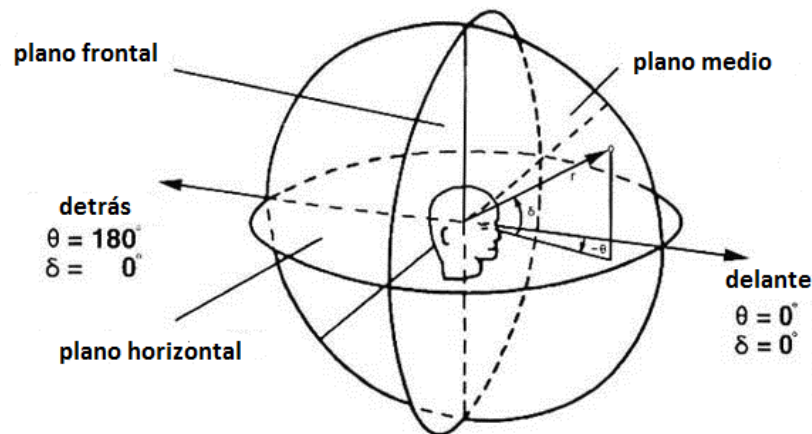


Figura 2. Detalle de la definición del sistema de coordenadas vertical-polar.

## II.3. DETERMINACIÓN DEL ÁNGULO LATERAL (AZIMUT)

La Teoría Duplex de John Strutt, más conocido como Lord Rayleigh, da cuenta de los principales factores que intervienen en la localización de fuentes [4]. Fue una teoría pionera en la investigación de la escucha espacial hace unos 100 años, pero que todavía se sigue utilizando. Principalmente se habla de la diferencia de tiempo interaural o ITD (*Interaural Time Difference*) y de la diferencia interaural de intensidad IID (*Interaural Intense Difference*). Estos son los aspectos más importantes para la localización de las fuentes tanto en el plano horizontal como en el vertical.

El hecho de tener dos oídos hace que existan unas pequeñas diferencias entre las señales que llegan a cada oído. Estas son las Diferencias Interaurales de Intensidad (IID) y las Diferencias



Interaurales de Tiempo (ITD) que se obtienen a partir de los distintos caminos que tiene que recorrer el sonido desde su origen (fuente sonora) hasta los tímpanos (receptor), pero también por la sombra acústica producida por la cabeza del individuo (difracción de la onda). Estos caminos producen diferentes atenuaciones que sufre cada señal a lo largo de sus recorridos, y diferentes distancias. Es por tanto, un proceso que se realiza con una escucha binaural. La cabeza es un obstáculo efectivo únicamente para las altas frecuencias (por encima de los 1000Hz), concretamente para aquellas que tienen longitudes de onda menores a la mitad del radio de la cabeza, por lo que se ha comprobado que no es posible localizar una fuente en el plano medio del espacio únicamente mediante Diferencias Interaurales de Intensidad (IID).

Por ello, Lord Rayleigh propone en su teoría las Diferencias Interaurales de Tiempo (ITD), que para sonidos senoidales sería la diferencia de fase o tiempo entre las ondas sonoras que llegan a cada uno de los oídos, debido a la orientación de la cabeza y la longitud de onda de los sonidos captados.

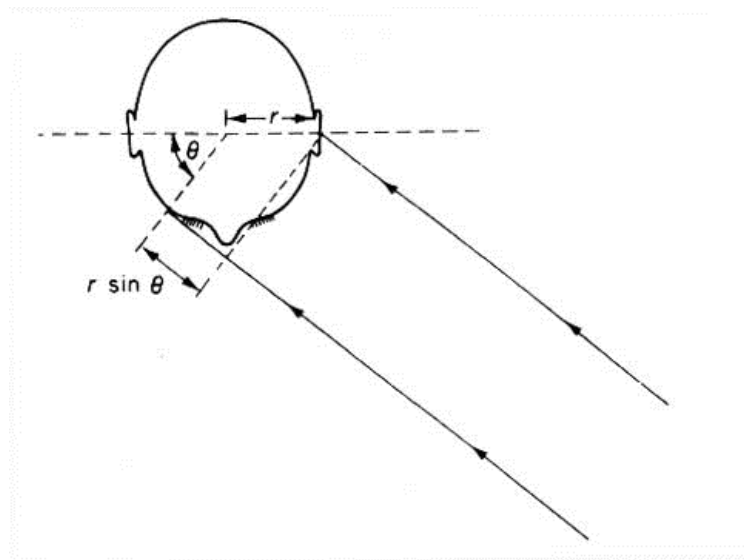


Figura 3. Diferencia Interaural de Tiempo en función del acimut.

A bajas frecuencias, esta diferencia supone un porcentaje muy pequeño de la longitud de onda, lo que hace muy difícil que una fracción tan pequeña se pueda detectar, mientras que para frecuencias más altas, 800 Hz por ejemplo, la separación de los oídos puede llegar a ser incluso de media longitud de onda, pudiendo dar lugar a ciclos enteros de diferencia de fase entre los sonidos que llegan a cada uno de los oídos, con lo que el sistema auditivo no tiene forma de determinar qué ciclo corresponde a cada uno de los oídos. Por tanto, para frecuencias altas, resulta ambiguo juzgar la dirección de la fuente sonora basándose sólo en la diferencia de fase entre el sonido percibido por los dos oídos y se puede concluir que la localización auditiva por

diferencia de fase (ITD) es definitiva en una banda de frecuencias comprendidas entre los 200 y los 800 Hz.

No se puede determinar la dirección de llegada de un sonido cotidiano con el alto grado de precisión que se obtienen en los cálculos con tonos puros, como hemos expuesto anteriormente. La naturaleza compleja de los sonidos, es debida a que el espectro de frecuencia se extiende sobre la mayor parte del rango audible, por lo que la diferencia entre niveles de intensidad y fases pueden ser utilizadas simultáneamente para localizar la procedencia del sonido.

Así que, para un caso práctico de localización de un sonido natural, para frecuencias altas (por encima de los 1000 Hz), la localización se debe fundamentalmente a las diferencias de intensidad, mientras que para las bajas frecuencias (menos de 800 Hz) la localización es debida a las diferencias de fase o tiempo de retardos en ambos oídos. En frecuencias medias existe una indeterminación que se resuelve tanto por medio de las diferencias de fase tanto como por las de nivel de intensidad para realizar la localización, es decir que utilizamos ambas de forma simultánea y combinada. Esta teoría fue denominada Teoría Duplex por Lord Rayleigh en 1907.

#### *II.4. DETERMINACIÓN DE LA ELEVACIÓN*

Mientras que la determinación del ángulo lateral, como ya se ha comentado, es un proceso binaural, la determinación de la elevación es monoaural y es debido a que nuestra oreja actúa como una antena acústica cuyas cavidades resonantes y geometría amplifican algunas frecuencias y atenúan otras. De hecho, su respuesta frecuencial depende de la dirección de llegada del sonido. Las partes del oído externo, como el pabellón de la oreja y la concha, e incluso las partes del ser humano cuerpo, como hombros, actuarán como filtros acústicos para cambiar la fase y la amplitud del sonido dependiendo de la dirección.

En la siguiente figura se aprecia la medida de la respuesta frecuencial del sonido de dos direcciones de llegada diferentes. En cada caso observamos que existen dos caminos desde la fuente hasta el canal auditivo: un camino directo y un camino más largo que sigue una reflexión en la oreja. Para frecuencias bajas, la oreja recibe más energía sonora y las señales de los dos caminos llegan esencialmente en fase. Sin embargo, para frecuencias elevadas, la señal retardada no está en fase con la señal directa y se producen interferencias destructivas. La mayor interferencia ocurre cuando la diferencia de longitud de los caminos ( $d$ ) es de media longitud de onda, es decir, cuando  $f = c/2d$ . En el ejemplo, esto produce una "hendidura" en el espectro recibido sobre los 10 kHz.

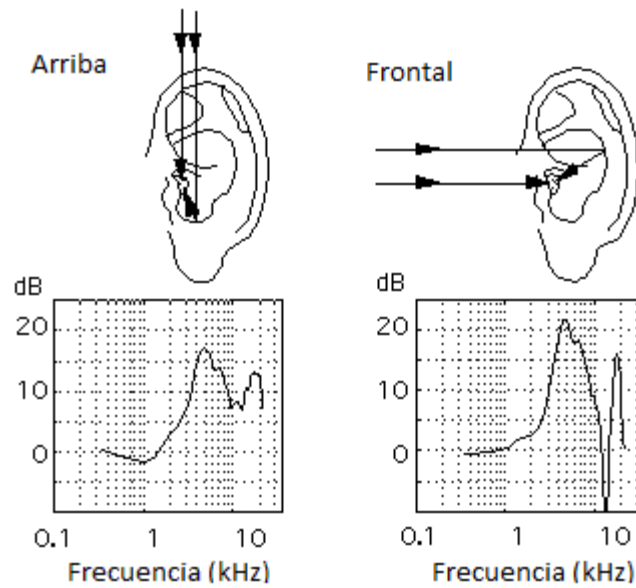


Figura 4. Filtrado de la señal emitida en el dominio frecuencial dependiente del ángulo de elevación.

La oreja recibe mucho mejor los sonidos que llegan de frente que los que llegan de otros ángulos como, en la figura, desde arriba. Por eso la hendidura resultante es mucho más pronunciada para fuentes frontales que fuentes situadas sobre la cabeza. Además, la diferencia de longitud entre los caminos varía con los cambios en el ángulo de elevación, por lo que la frecuencia de la hendidura varía con la elevación. Todavía hay debates sobre qué características son las más influyentes, pero lo que sí está bien aceptado por la comunidad científica es que la oreja es la responsable de la detección del ángulo de elevación.

### II.5. AMBIGÜEDAD EN LA LOCALIZACIÓN

Existen casos de ambigüedad en la localización de la fuente, como por ejemplo en casos donde los sonidos proceden de dos fuentes situadas de forma simétrica y a la misma distancia de la cabeza. En este caso, los sonidos llegan al oído izquierdo con la misma diferencia de tiempo con respecto al derecho que de modo contrario. Afortunadamente la forma asimétrica del pabellón auditivo ayuda a aclarar cualquier tipo de duda porque el sonido se modifica sustancialmente antes de llegar al tímpano dependiendo de la dirección de llegada. Otro caso se da en recintos cerrados, donde existen reflexiones sonoras ya que el oyente recibe diferentes señales, unas provenientes directamente de la fuente sonora y otras que llegan después de reflejarse, una o más veces, en las superficies interiores del recinto o en los objetos que contiene el mismo. Estas señales alcanzan los oídos del oyente con diversos retardos de tiempo y desde direcciones diferentes. Como el camino más corto entre dos puntos es la línea recta, el sonido que llega primero es el directo, que es el utilizado para localizar la fuente sonora por medio de la escucha

binaural [5] [6]. Posteriormente llegan las reflexiones, que han perdido parte de su potencia en los choques con las superficies reflejantes, por lo que contienen menos energía que el sonido directo. Estos sonidos reflejados proporcionan el ambiente acústico o reverberación. Hay varios tipos de efectos temporales, entre ellos dos casos extremos: los sonidos reflejados que llegan inmediatamente después del directo y que contribuyen a enriquecer la tonalidad del sonido original, y los sonidos que llegan mucho más tarde y que se perciben como ecos.

Entre estas dos situaciones, existen una gran cantidad de posibles casos intermedios que fomentan la ambigüedad en la localización. Cuando el sonido es continuo, el oído recibe una información directa y otra reflejada, que son idénticas en tono y cuyas intensidades varían dependiendo de las condiciones del recinto. Por esta razón es muy difícil localizar en salas normales tonos puros cuando estos se están produciendo. En cambio, si estos tonos se interrumpen momentáneamente puede determinarse su origen, ya que los primeros ciclos llegan con un cierto adelanto al oído con respecto a las reflexiones, que inevitablemente se retrasarán [6] [7].

Otro posible caso de ambigüedad se produce cuando la fuente sonora está detrás o delante del oyente. En este caso, los sonidos llegan al oyente con igual intensidad y fase a ambos oídos y es imposible discernir su origen. Se podría solucionar el problema moviendo la cabeza con el objetivo de establecer unas diferencias temporales que nos ayuden en la localización, aunque la escucha binaural mediante auriculares no permita dicho movimiento, porque, todo el espacio acústico se movería con el movimiento de la cabeza. Además se ha comprobado también que el Efecto de Precedencia o Efecto Haas, influye drásticamente en la localización del sonido que se puede generar cuando un oyente que “ve” dos altavoces situados a la misma distancia bajo un ángulo de 45°, que percibe dos emisiones idénticas (cada una de ella procedente de un altavoz), que no se diferencian entre ellas más que por el intervalo temporal que separa su llegada al oído [5] [7]. Los experimentos, además, han llevado a sugerir que el efecto de precedencia tiene su explicación en procesos cognitivos que se desarrollan en niveles relativamente altos [5] [8].

## *II.6. HEAD-RELATED TRANSFER FUNCTIONS (HRTF)*

Mientras las relaciones entre los ITD e IID y la localización espacial eran razonablemente predecibles, la comunidad científica encontró problemas para relacionar la percepción espacial los sonidos con su variación en el contenido espectral mediante expresiones matemáticas. Por ello, el primer paso hacia la comprensión de los cambios de contenido espectral con la escucha direccional fue la realización de modelos físicos [9], además de medidas y experimentos empíricos [10], más recientemente simulaciones computacionales [11], para la obtención de la variación

del contenido espectral dependiente de la dirección obtenida directamente de la oreja física de una persona.

Así, estas medidas fueron denominadas *Head-Related Transfer Functions* (HRTF) o Funciones de Transferencia Relativas a la Cabeza y resumían el filtrado acústico dependiente de la dirección producido por el torso, la cabeza y pinna sobre un sonido emitido en campo libre. De esta manera, los científicos esperaron primero obtener la respuesta en frecuencia de un oído y así analizar las variaciones de la percepción respecto a la dirección de llegada del sonido. Formalmente, las HRTF se definen como la respuesta en frecuencia en campo lejano de un oído específico (izquierdo o derecho) de cada individuo, medido desde un punto concreto del campo libre a un punto concreto del oído del individuo, normalmente para un radio fijado de distancia desde la cabeza del oyente. Estas medidas se realizan para diferentes variaciones del punto de emisión en grados o radianes, del plano horizontal (acimut) y el plano frontal (elevación) [10].

Sin embargo, el efecto de la distancia de fuente se ha omitido en la mayoría de estudios de HRTF. En el campo lejano los efectos con la distancia son mínimos [12][13], pero conforme la fuente se va acercando al oyente (alrededor de 1m.) las curvas interaurales, sobre todo las ILD, se deberían de representar como funciones de la distancia. En la literatura hay sorprendentemente muy poca información sobre el comportamiento de las HRTF en el campo cercano [12][14], aunque sí que se encuentra algún estudio al respecto que proponen modelos para la distancia [4]. En nuestro caso no vamos a utilizar estos modelos, ya que vamos a simular las simulaciones dentro de una sala. Así, el efecto de la distancia lo recrearemos utilizando el efecto de la reverberación de la propia sala.

Por lo tanto, la HRTF es una función de cuatro variables (si no contamos la distancia de la cabeza a la fuente, tres): tres coordenadas espaciales y la frecuencia. En coordenadas esféricas, para distancias mayores de un metro, se dice que la fuente está en el campo lejano y la HRTF decae inversamente con la distancia. La mayoría de las medidas de la HRTF se realizan en el campo lejano, lo que reduce la HRTF a una función del acimut, la elevación y la frecuencia [15]. Cabe mencionar que mediante la Transformada de Fourier Inversa se consigue la expresión análoga en el dominio temporal denominada Respuesta al Impulso Relativa a la Cabeza o *Head-Related Impulse Response* (HRIR).

Las HRIR son usualmente definidas como filtros de Respuesta al Impulso Finita o FIR y suman la información de ambos sistemas definidos por Lord Rayleigh en la Teoría Dúplex, IID e ITD de forma que los retardos de tiempo (o la información de ITD) son codificados en el

espectro en la fase del filtro y la información de las diferencias de nivel es codificada en el módulo en función de la frecuencia del filtro [16].

### III.EXTERNALIZACIÓN DEL SONIDO

#### III.1. TEORÍA PARA LA EXTERNALIZACIÓN DEL SONIDO

Cuando tratamos de localizar una fuente sonora, somos mejores estimando el azimut, un poco peores estimando la elevación y poco hábiles determinando la distancia a la fuente. De la misma manera, comprendemos muy bien los mecanismos para determinar el ángulo lateral, un poco peor los que estiman la elevación y aún peor para entender el mecanismo que nos dice a que distancia está la fuente.

Generalmente se tienen en cuenta los siguientes mecanismos para determinar la distancia de la fuente sonora al oyente: *loudness* o amplitud del sonido, movimiento paralelo, exceso en las IID (*Interaural Intense Difference*) y la razón entre el sonido directo y el sonido reverberante.

Los principios físicos del mecanismo de detección por *loudness* se deben obviamente por el hecho de que la energía sonora capturada proveniente directamente de la fuente decae inversamente con el cuadrado de la distancia. Cuando una fuente de energía constante se aproxima al oyente, la amplitud del sonido se incrementa. Es igualmente obvio que la energía recibida es proporcional a la energía emitida por la fuente, y que no puede haber una relación uno-a-uno entre el *loudness* y la distancia. Por ejemplo, reproducir un sonido a un volumen bajo no produce por sí mismo la sensación de que la fuente esté muy lejos. Para usar el *loudness* como medida de la distancia, debemos saber también algo acerca de las características de la fuente. En el caso de la voz humana, cada uno de nosotros sabemos la diferencia entre la calidad del sonido asociada a un susurro, a una conversación normal o a un grito sin tener en cuenta el nivel sonoro. La combinación del *loudness* y el conocimiento de la fuente nos proporcionan una información útil para juzgar la distancia.

El movimiento paralelo se refiere al hecho de que si el oyente mueve su cabeza, el cambio de azimut puede depender de la distancia. Para fuentes muy próximas, un movimiento pequeño puede provocar un gran cambio en el ángulo lateral, mientras que para fuentes lejanas esencialmente no hay cambios de azimut. Esta información es también muy útil a la hora de identificar la distancia de la fuente sonora.

Además, cuando una fuente se aproxima mucho a la cabeza, las IID (*Interaural Intense Difference*) pueden aumentar. Este aumento se hace patente para distancias por debajo de un metro, aproximadamente. Un caso extremo se da cuando un insecto zumba en un oído o cuando alguien susurra en un oído. En general, sonidos escuchados en un solo oído son amenazadores e incómodos de oír.

El último mecanismo utilizado para identificar la distancia de la fuente sonora es la razón entre el sonido directo y el reverberante. Como ya se mencionó antes, la energía recibida de una fuente sonora decae inversamente con el cuadrado de la distancia. De todos modos, en habitaciones ordinarias, el sonido es reflejado y dispersado muchas veces por las superficies del entorno, y la energía reverberante que llega a los oídos no cambia mucho con la distancia entre la fuente y el oyente. Por ello, esta razón es el mejor mecanismo para determinar la distancia. Para distancias cortas, esta razón es muy grande, mientras que para distancias largas es pequeña.

### *III.2. MODELIZACIÓN CON LA DISTANCIA*

Escuchar la música a través de auriculares con reproductores multimedia portátiles se ha convertido en algo muy común últimamente. Si el sonido no ha sido postprocesado correctamente con algoritmos específicos diseñados para auriculares, es probable que éste se localice dentro de la cabeza, creando una imagen lateralizada entre los oídos que puede provocar una sensación poco natural para el oyente.

Los sistemas comunes de mejora de sonido para auriculares se pueden clasificar en tres categorías: algoritmos de procesamiento espacial simplificados, algoritmos de procesamiento binaural basados en HRTF y algoritmos de posicionamiento de fuentes virtuales. En este trabajo todos estos tipos son investigados y se combinan entre sí para encontrar un único método que produce el resultado de audio binaural reproducido teniendo la mejor localización de la fuente ‘externalizada’ usando auriculares.

Las HRTF (*Head-Related Transfer Functions*) son medidas basadas en la aproximación del efecto que la oreja humana añade al sonido recibido. Algunos sistemas de externalización del sonido basados en HRTF han sido propuestos previamente. Estos se basan en la dependencia de cada pabellón auditivo independiente deformando el timbre del sonido. También se han propuesto procesados espaciales más ligeros. Sus ventajas son la reducción del coste computacional y la independencia con el oyente. Sin embargo, estos procesados no son siempre óptimos en la generación del efecto de externalización del sonido.

La HRTF varía con la distancia al igual que con el azimut o la elevación. Debido a la absorción del aire, ciertas frecuencias se atenúan más que otras, ya que las ondas sonoras viajan a través del aire. La HRTF de la oreja en la que se produce una sombra acústica, difiere radicalmente de la de la oreja sin sombra. Cuanto mayor es el ángulo de azimut, más sombra acústica va a producir la cabeza. La difusión de las ondas sonoras hace que las bajas frecuencias estén menos expuestas a la alternancia de las HRTF. Todas estas dependencias y sus relaciones con HRTF pueden ser explotadas en el proceso de diseño de un algoritmo DSP para desviar el sistema auditivo.

Además de estos sistemas, debemos introducir el efecto que produce la sala en la que se desarrolla el proceso. Para ello se calcula la RIR (*Room Impulse Response*) que es la respuesta al impulso de la sala. Esa señal caracteriza las propiedades de la sala y como modifica cualquier señal de audio, a partir de las posiciones de la fuente emisora y el receptor.

Existen tres teorías diferentes que estudian la acústica de salas. Son la teoría estadística, la teoría geométrica y la teoría ondulatoria. Por eso hay también diferentes métodos de calcular la RIR atendiendo a estas teorías.

La Teoría Estadística estudia la energía acústica presente en un recinto sin tener en cuenta el carácter ondulatorio del campo sonoro. Establece que todas las partículas sonoras son idénticas en tamaño y energía. Cuando una fuente sonora irradia energía acústica en un recinto cerrado, el campo acústico resultante empieza a propagarse libremente. Cuando las ondas acústicas inciden en los cerramientos, éstas empiezan a sufrir reflexiones. En el momento en que el recinto absorbe la misma energía por unidad de tiempo que la emitida por la fuente, la energía presente en la sala permanece estacionaria a un cierto nivel; en este momento la energía media en la sala permanece constante.

La teoría geométrica estudia el campo sonoro mediante un análisis acústico basado en la hipótesis de reflexiones especulares. Para ello se introduce el concepto de rayo sonoro y su estudio se asemeja al estudio de la física óptica. Las reflexiones del rayo sonoro cumplen el principio de Fermat que verifica la ley de la reflexión.

La teoría ondulatoria tiene en cuenta el carácter ondulatorio del sonido. Se basa en la resolución de la ecuación diferencial de la onda acústica en la sala. En la resolución de esta ecuación se tiene en cuenta las condiciones del entorno como puede ser los cerramientos, etc.



La teoría ondulatoria ofrece resultados precisos pero tienen un elevado coste computacional. Por otro lado la teoría estadística no ofrecen resultados precisos mientras que las técnicas geométricas se pueden dividir en dos categorías: trazado de rayos imágenes especulares. Hemos seleccionado estas últimas por su compromiso entre precisión coste computacional.

#### **IV. SELECCIÓN DEL MODELO Y LENGUAJE DE DESARROLLO**

En este capítulo, valoraremos la plataforma para desarrollar e implementar los algoritmos descritos anteriores para obtener el efecto de la auralización del sonido. Como hemos explicado, la auralización se compone fundamentalmente de dos efectos: el efecto de holofonía el cual se implementará mediante el filtrado de la señal acústica con las **HRTF** s, y el efecto de la externalización, que lo implementaremos mediante el filtrado de la señal con la **RIR** . A continuación se describe y se justifica la elección del modelo y del lenguaje de desarrollo elegidos para esta implementación.

##### *IV.1. ELECCIÓN DEL LENGUAJE DE DESARROLLO*

La selección del lenguaje de desarrollo puede limitar en algunos aspectos, y por eso debemos ser cuidadosos. Existen muchos lenguajes de programación, y seguramente con más de uno se puede desarrollar la misma aplicación, pero se han de tener en cuenta algunos aspectos a la hora de tomar la decisión, como coste computacional, simplicidad en la programación, acceso a la interfaz y también conocimientos previos.

Debemos buscar un lenguaje de desarrollo adecuado, potente, sencillo y capaz de satisfacer las necesidades del trabajo a realizar. Entre las múltiples opciones, destacamos MatLab ya que nos permite realizar la mayoría de las aplicaciones que pretendemos programar con bastante simplicidad al disponer de un gran número de funciones predefinidas para el procesado de señales de audio. Por otro lado, MatLab es el lenguaje que se ha utilizado a lo largo del desarrollo del Máster, con lo que partimos de unos conocimientos previos técnicos que nos facilitarán el trabajo. Además, la universidad facilita su uso a partir de su servidor de licencias.

Destacamos algunas de las funciones de MatLab que utilizamos para desarrollar la aplicación como: *filter*, *conv* y *audio*. Estas funciones nos facilitan el trabajo con señales de audio y su reproducción en tiempo real. Las funciones *filter* y *conv* sirven para realizar el filtrado de señales. La diferencia es que la primera realiza el filtrado por bloques, es decir, nos da un poco más de margen ya que podemos cambiar los coeficientes del filtro conforme estamos

filtrando la señal; mientras que la segunda realiza el filtrado de toda la señal. Por otro lado la función *audio* nos va a permitir reproducir la señal de audio.

#### IV.2. IMPLEMENTACIÓN DE LA HOLOFONÍA: HRTF

Como ya se ha comentado, la HRTF modela la posición de la fuente alrededor de la cabeza, por ello necesitamos una base de datos de HRTF que cubra la mayoría de ángulos posibles.

Hemos seleccionado la base de datos proporcionada por Oliver Warusfel y el Equipo de Acústica de Salas del Instituto de Investigación y Coordinación de la Música/Acústica de París (Institut de Recherche et Coordination Acoustique/Musique, IRCAM) [17] por varios motivos. Esta base de datos dispone de una gran amplitud de ángulos en azimut y elevación. La cobertura es de 135° (desde -45° hasta 90°) para la elevación, con una discretización cada 15°. Los saltos de rotación o acimut varían según el ángulo de elevación desde 24 medidas (una medida cada 15° a lo largo de 360°) a una única en +90° de elevación. En total disponemos de 187 señales de HRTF. Además las señales están dispuestas para ser utilizadas directamente por MatLab, que es el lenguaje en el que vamos a desarrollar la aplicación. Después de escuchar algunas señales de audio grabadas en salas anecoicas y combinadas con estas HRTF por varios sujetos, decidimos que estas HRTF son las que vamos a utilizar durante el transcurso del trabajo.

Este instituto, ofrece un conjunto de HRIRs medidas en cámara anecoica, en los oídos de un conjunto de 51 sujetos. Se decide tomar esta biblioteca de todas las existentes en la red por la simplicidad que supone tener los ficheros en el formato de datos de MatLab para su posterior uso. La única desventaja de esta biblioteca es que no contiene la respuesta de un maniquí especialmente diseñado para este tipo de medidas por lo que se ha optado por utilizar la HRTF que más se ajustaba al autor.

<b>Elevación (Grados)</b>	<b>Incremento de acimut (Grados)</b>	<b>Puntos por elevación</b>
-45	15	24
-30	15	24
-15	15	24
0	15	24
15	15	24
30	15	24
45	15	24
60	30	12
75	60	6
90	360	1

Tabla 1. Resumen del protocolo de medida de las HRTF.

Además de la base de datos, el IRCAM ofrece ciertos códigos de ayuda implementados sobre MatLab como por ejemplo las sentencias necesarias para la llamada a las HRIRs o el código necesario para compensar el retardo de propagación y partir la función de transferencia en filtros de fase mínima y componentes paso todo.

Las medidas han sido tomadas en una cámara anecoica. Unos pequeños micrófonos han sido colocados en la oreja de los sujetos mediante el método *blocking-meatus*, lo que supone que la parte frontal de la cápsula se sitúa en la entrada del canal auditivo (con ayuda de un molde de silicona realizado para cada sujeto) bloqueando la entrada de sonido en el mismo para evitar resonancias, siendo esta la manera óptima de hacer estas medidas según distintos autores [18][19].

Con estas señales de HRTF se crea el efecto de auralización del sonido, en el que la fuente parece llegar desde un punto en el espacio, aunque estemos escuchando el sonido a través de unos auriculares. Las HRTF tendrán que ir variando en función de la orientación que nos proporciona el posicionador o Tracker. Al adaptar las HRTF en función de la posición del usuario, conseguimos el efecto de espacialidad deseado, a pesar de que el sonido venga a través de unos auriculares. Para realizar este filtrado utilizaremos la función *filter* de MatLab que, como ya hemos explicado, nos permite realizar un filtrado por bloques, pudiendo ir cambiando el tipo de filtro durante el proceso. Esto es justo lo que queremos, ya que durante el filtrado tenemos que ir cambiando las HRTF según va cambiando la orientación del oyente.

A través de las coordenadas y la orientación que nos proporciona el posicionador o Tracker, obtenemos los datos necesarios para buscar en la base de datos las señales correspondientes de HRTF que han de ser utilizadas para realizar el filtrado con la señal de entrada consiguiendo así el efecto de espacialidad deseado. Para realizar este filtrado utilizamos la función *filter* de Matab.

Debido a que no disponemos de una base de datos continua para todos los ángulos en azimut y elevación, debemos de tener especial cuidado a la hora de elegir las señales de HRTF para realizar el filtrado. Para la aplicación vamos a utilizar la señal de HRTF disponible en la base de datos más cercana a la aproximación de los datos de posicionamiento y orientación dados por el Tracker para cada instante. Esto en algunos casos puede generar una apreciación de saltos, creando una sensación extraña. Dejando como trabajo futuro la interpolación para los diferentes ángulos.

### IV.3. IMPLEMENTACIÓN DE LA EXTERNALIZACIÓN (REVERBERACIÓN): RIR

Tenemos que modelizar el efecto de la reverberación de la sala en la señal de entrada. Existen tres teorías diferentes que estudian la acústica de salas. Son la teoría estadística, la teoría geométrica y la teoría ondulatoria. Por su compromiso entre coste computacional y precisión vamos a utilizar el método de las imágenes, que se basa en la teoría geométrica. La aplicación de éste método esta limitada al estudio de salas de paredes planas y con modelos de reflexiones especulares geométricas, como por ejemplo, al estudio de salas rectangulares, en las que el cálculo de las posiciones de las fuentes imagen es muy sencillo. Esta premisa se adapta a nuestro modelo, ya que nuestra sala será una pequeña habitación rectangular.

Según este modelo, una fuente sonora situada frente a una pared rígida genera una onda reflejada cuya dirección e intensidad sonora son idénticas a las de la onda directa de una fuente virtual situada al otro lado de la pared, y que empezó a emitir al mismo tiempo que la original. Por supuesto, en los correspondientes cálculos se debe tener en cuenta la absorción de la pared, disminuyendo el valor de la intensidad sonora en el factor correspondiente.

De esta manera, y siguiendo el mismo proceso para todas las superficies, se pueden encontrar todas las reflexiones de primer orden para cualquier receptor en la sala. Este proceso se repite para cada una de las fuentes imagen obtenidas, dando lugar a las fuentes imagen responsables de las reflexiones de segundo orden, y así sucesivamente.

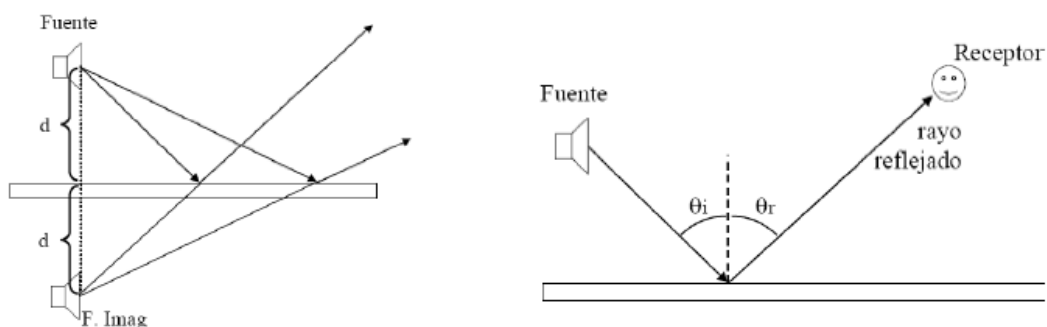


Figura 5. Reflexión especular de los rayos, según el principio de Fermat.

Se ha realizado también una búsqueda y estudio de diferentes algoritmos dedicados al efecto de la reverberación de salas.

Stephen G. McGovern utiliza un simple algoritmo basado en el método de las imágenes para calcular la respuesta al impulso de la sala (RIR). A partir del orden de fuentes imagen que se desean calcular, el coeficiente de absorción global de los cerramientos y las dimensiones de

la sala (alto, ancho y largo), se calcula la respuesta al impulso de la sala, para una posición fija de fuente sonora y receptor. Así, para añadir el efecto de la sala a la señal de entrada, solo tendremos que convolucionarla con la  $h_{\text{sala}}$  que obtendremos gracias a este algoritmo. A continuación explicamos más detalladamente cómo funciona este algoritmo.

En primer lugar se calculan los ecos individuales de cada fuente imagen que en conjunto producen la reverberación. En segundo lugar se calcula la respuesta al impulso unitaria para cada eco con el retardo de tiempo adecuado, para que sea escuchado en una posición particular de la sala. En tercer lugar se calcula la magnitud de la respuesta impulsiva unitaria de cada eco. Lo que es más o menos vamos a hacer es encontrar el tiempo y la magnitud de cada eco que se escucha desde una posición particular de la habitación. Finalmente toda esta información se juntará en una función unidimensional de tiempo. Esta función será nuestra respuesta impulso de la sala. En la función de respuesta al impulso, el tiempo se podrá discretizar. El tiempo discreto nos permitirá utilizarlo como un filtro FIR (*Finite Impulse Response*) para la simulación de la reverberación.

El mayor problema de este algoritmo, es su gran coste computacional a la hora de obtener la  $h_{\text{sala}}$  ya que utiliza un filtro FIR de con un gran número de coeficientes para realizar dicho cálculo. Por ello empezamos a investigar otros métodos alternativos de cálculo de la reverberación de salas mediante filtros IIR con ánimo de reducir ese coste computacional.

Lamentablemente, tras hacer diversas pruebas nos damos cuenta de que ningún filtro IIR simple (con pocos coeficientes) puede modelar esta RIR. Sin embargo, la reverberación de la sala podría dividirse en dos: la reverberación temprana, y la reverberación tardía. La reverberación temprana asumimos que es difícil de modelar mediante un filtro IIR ya que necesitaríamos un filtro de un orden muy elevado, mientras que la reverberación tardía pensamos en un principio que sí sería posible. A continuación se muestra una gráfica que representa el error acumulado de la reverberación temprana (rojo) y el error acumulado de la reverberación tardía (azul) que resultan de utilizar un filtro IIR de 5 coeficientes en el numerador, y 6 coeficientes en el denominador.

Se observa que el error en la reverberación temprana aumenta mucho más rápidamente en comparación con el error de la reverberación tardía. Sin embargo, el error de la reverberación tardía también termina aumentando bastante.

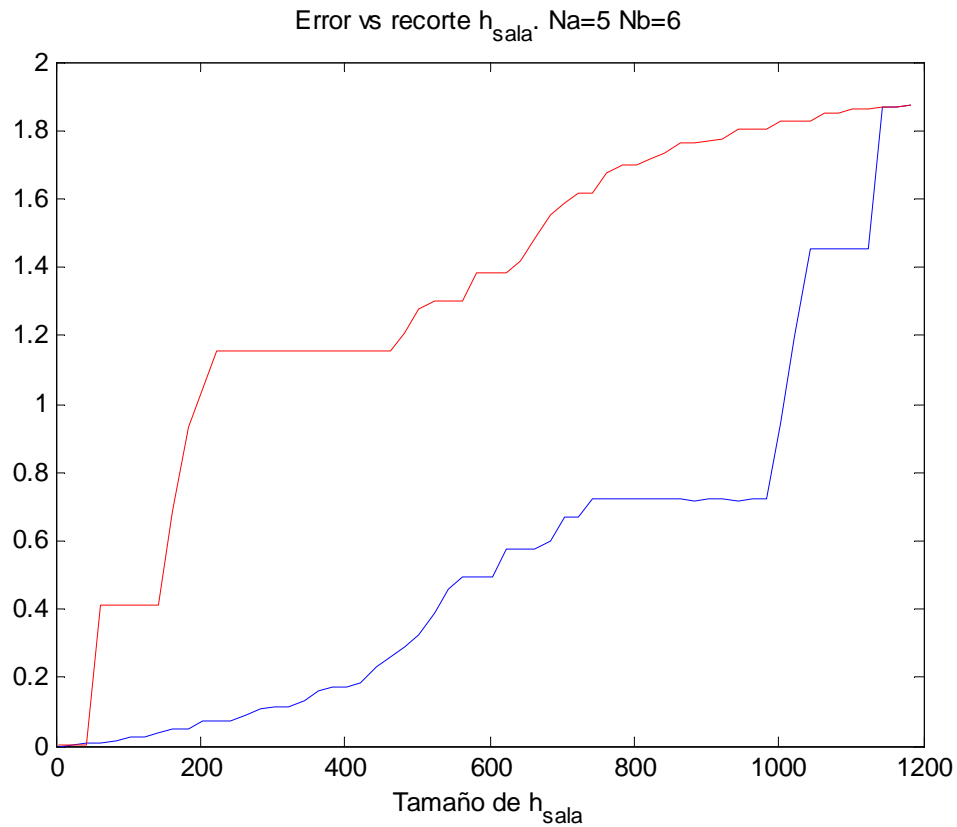


Figura 6. Error acumulado reverberación temprana (rojo) y reverberación tardía (azul).

Entonces pensamos en utilizar dos filtros, un filtro FIR que caracterice la reverberación temprana, y un filtro IIR que se encargue de caracterizar la reverberación tardía, reduciendo así en gran medida los coeficientes de el filtro FIR original.

Aparentemente no aparece un modelo subyacente q permita reducir el número de coeficientes del filtro con el fin de acortar la respuesta de la sala. Cuanto más pequeño es el número de coeficientes, más rápido crece el error acumulado. Por eso no nos conviene usar un filtro IIR si para ello tenemos que utilizar un número muy elevado de coeficientes. A continuación se puede ver una gráfica que nos muestra el error acumulado del filtro IIR:

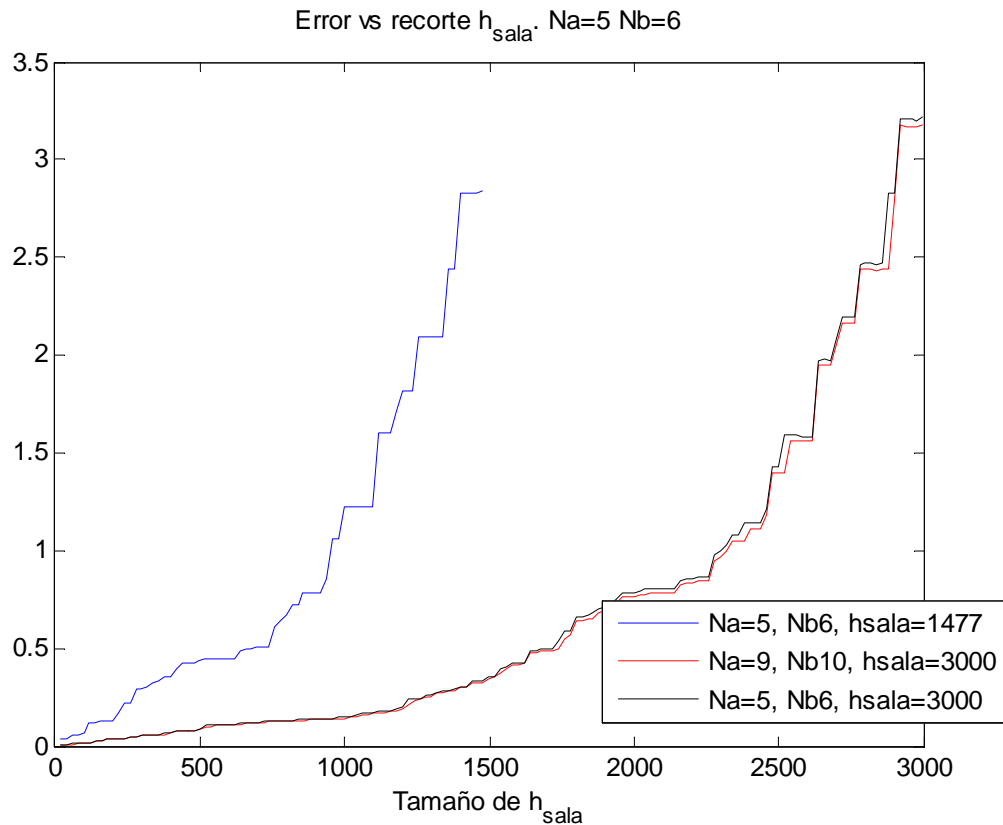


Figura 7. Error acumulado reverberación tardía con diferentes filtros IIR.

Como no conseguimos ningún resultado convincente, vamos a utilizar el algoritmo original de McGovern para realizar los cálculos. Pero como ya hemos mencionado, el coste computacional de este algoritmo es bastante elevado. Y más aún a la hora de ser utilizado en tiempo real. Necesitamos una  $h_{sala}$  adecuada a la posición del receptor para cada instante de tiempo. El límite de coste de tiempo para realizar los cálculos viene dado por el intervalo temporal que vayamos a utilizar. Desde luego el intervalo a de ser lo suficientemente pequeño para que el oyente no note saltos en la señal de salida, empeorando la calidad de la grabación.

Definitivamente es prácticamente imposible reproducir señales en tiempo real teniendo que calcular esta  $h_{sala}$  para cada muestra temporal. Por eso, la solución que vamos a practicar es crearnos una base de datos de  $h_{sala}$  para un mallado de posiciones de receptor, y por el cual se va a poder mover nuestro oyente. De esta manera, sólo se realiza una primera simulación con un elevado coste computacional, para obtener así una  $h_{sala}$  por cada punto de la malla predefinida inicialmente. Después durante la reproducción en tiempo real de la simulación, sólo tendremos que seleccionar la señal  $h_{sala}$  correspondiente según los datos del posicionador.

#### IV.4. INTEGRACIÓN DE LAS HRTF Y RIR

Una vez elegida la base de datos de HRTF y seleccionado un algoritmo que nos genera una pequeña base de datos con diferentes respuestas al impulso de la sala con respecto a la posición del receptor, sólo tenemos que agruparlo todo para obtener el resultado final. Vamos a crear un programa que a partir de los datos de posición (x,y) y orientación ( $\theta, \tau$ ) nos devuelva las HRTF correspondiente (dependiendo de la orientación del receptor con respecto de la fuente) y la respuesta al impulso de la sala correspondiente (dependiendo de la posición del receptor con respecto a la fuente).

Antes de empezar a desarrollar la aplicación, nos encontramos ante dos posibles configuraciones o modelos válidos para la integración de las HRTF con la respuesta al impulso de la sala. Las configuraciones son la aditiva (suma la señal de HRTF con la RIR) y la convolutiva (convoluciona la señal de HRTF y RIR).

El modelo convolutivo realiza la convolución de la señal de entrada con la  $h_{sala}$  y después de nuevo la convolución con la HRTF de cada canal correspondiente. Así la  $h_{total}$  del canal izquierdo quedará de la siguiente manera:

$$h_{total, \text{left}}(n) = h_{sala}(\bar{x}_{tx}, \bar{x}_{rx}, n) * h_{HRTF, \text{left}}(\Delta\bar{\varphi}, n)$$

Y su diagrama de bloques quedaría de la siguiente manera:

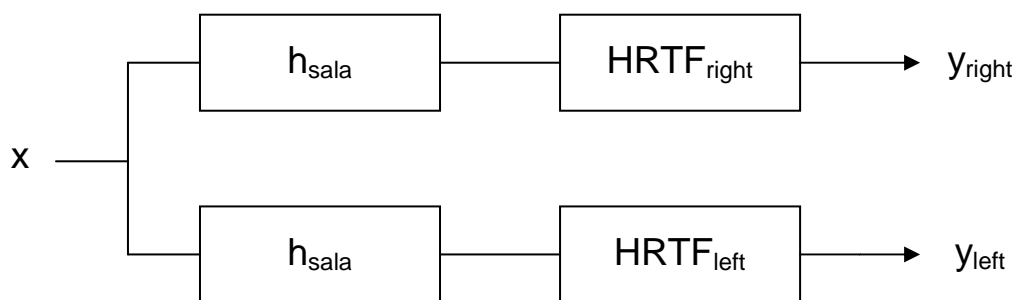


Figura 8. Modelo convolutivo de integración de  $h_{sala}$  y HRTF.

Por otro lado el modelo aditivo realiza la convolución de la señal de entrada con la  $h_{sala}$  y la convolución de la señal de entrada con la HRTF de cada canal correspondiente para después sumar los resultados. Así la  $h_{total}$  del canal izquierdo quedará de la siguiente manera:



$$h_{total,left}(n) = h_{sala}(\bar{x}_{tx}, \bar{x}_{rx}, n) + h_{HRTF,left}(\Delta\bar{\varphi}, n)$$

Y su diagrama de bloques correspondiente de la siguiente manera:

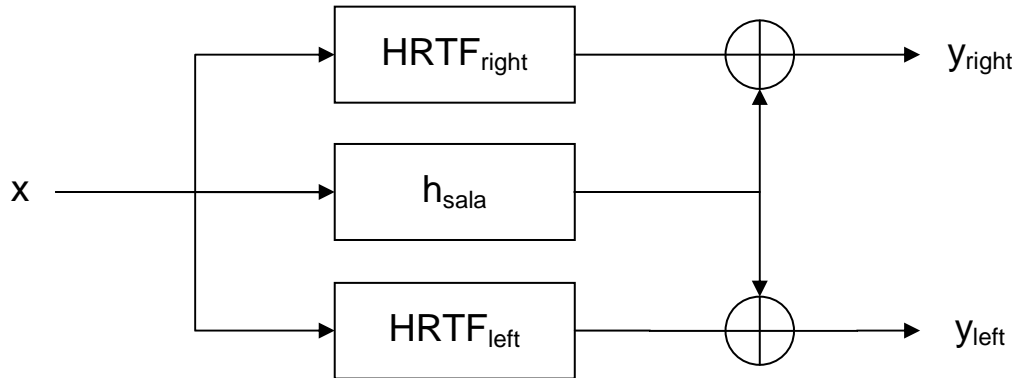


Figura 9. Modelo aditivo de integración de  $h_{sala}$  y HRTF.

Creamos el fichero *main.m* que se encarga de crear dos señal de salida, una por cada modelo, para poder así después ser comparados. Básicamente este fichero realiza la búsqueda en las bases de datos de las señales de HRTF y RIR correspondientes en cada instante de tiempo para una posición y orientación dadas, y las integra mediante el modelo convolutivo y el aditivo, creando así dos señales de salida listas para ser reproducidas. Los datos de posición y orientación los creamos primeramente mediante un vector de movimiento simulado, para más adelante ser leídos a partir de los datos proporcionados por *tracker* (“5.2 ELECCIÓN DEL TRACKER”).

Para seleccionar un modelo, se realiza una campaña experimental (“6. CAMPAÑA DE MEDIDAS campaña de medidas”) en el que se generan señales a partir de ambos métodos, y el que mejor respuesta tenga es el método que vamos a utilizar en el desarrollo de la aplicación.

## V. GENERACIÓN DEL EFECTO DE AURALIZACIÓN EN TIEMPO REAL

### V.1. INTRODUCCIÓN

Ya tenemos todos los módulos para crear el efecto de auralización (efecto 3D) tanto estático como dinámico (desplazamiento o no entre la fuente y el receptor) a partir de una señal de audio anecoica, pero en diferido, es decir, sin capacidad de adaptar estos efectos en tiempo real. Ahora lo que se pretende es conseguir que este mismo programa sea capaz de funcionar en tiempo real adaptando los efectos de tridimensionalidad conforme la señal se vaya reproduciendo y

conforme el oyente se vaya moviendo. Para conseguirlo, tenemos que realizar algún cambio en la configuración que tenemos hasta ahora. Lo que haremos fundamentalmente será sustituir el vector de movimiento que simulaba el desplazamiento, por una lectura de la posición y orientación del receptor, obtenida a partir de un *Tracker* en tiempo real. A partir de esta posición y orientación, se realizará la búsqueda de las señales de HRTF y RIR para filtrar la señal y adaptar la sensación de tridimensionalidad al movimiento del oyente capturado por el *tracker*.

En este capítulo describiremos los pasos que se han seguido desde la elección del *tracker* hasta su integración en la aplicación, hasta los resultados obtenidos.

## V.2. ELECCIÓN DEL TRACKER

Los *trackers* o posicionadores son dispositivos que miden el movimiento de "objetos" por ejemplo el de la muñeca o el de la cabeza del usuario con respecto a un sistema de referencia. Cuando este sistema es fijo, se denominan *trackers* absolutos, y cuando no es fijo, se denominan inerciales.

Existen diferentes características de los *trackers* que hay que tener en cuenta a la hora de la elección del mismo. A continuación se enumeran las principales características:

- Tasa de medida: lecturas por segundo que el *tracker* proporciona al ordenador.
- Latencia: retraso entre un cambio en la posición y orientación del objeto y el momento en el que se informa de ese cambio al ordenador.
- Precisión: diferencia entre la posición real del objeto y la que proporciona el *tracker*. Se degrada con la distancia.
- Volumen de trabajo: volumen dentro del cual el *tracker* puede medir y proporcionar dicha precisión.
- Resolución: cambio mínimo en la posición del objeto, que el *tracker* es capaz de detectar.
- Repetibilidad: medidas repetidas para una posición quieta del objeto.
- Ruido: cambios en la salida proporcionada por el *tracker* cuando el sujeto esta quieto.

Existen diferentes tipos de *trackers* atendiendo a la tecnología que utilizan para detectar el movimiento como son los *trackers* mecánicos, magnéticos, ópticos, de ultrasonidos, y los inerciales. A continuación se describen las principales características de cada tipo, así como sus ventajas y desventajas.

Los *trackers* mecánicos miden la posición y orientación de un objeto que está unido al extremo de un brazo mecánico. Este brazo está anclado a un punto fijo y compuesto de diferentes secciones que se pueden rotar y mover en las articulaciones. Los movimientos se miden mediante potenciómetros o codificadores y se usan para calcular la posición y orientación relativas al punto fijo de referencia. Estos dispositivos tienen una latencia muy baja por lo que son muy rápidos. Además son inmunes a interferencias de campos magnéticos. Pueden llegar a ser pesados si se usa sobre el cuerpo del usuario. Tienen poca libertad, es decir, un volumen de trabajo pequeño, y en general son poco usados.

Los *trackers* magnéticos son dispositivo de medida sin contacto que utiliza un campo magnético producido por un transmisor fijo para determinar la posición en tiempo real de un elemento receptor móvil (sobre el objeto). Estos dispositivos usan campos magnéticos de baja frecuencia para medir la posición. Estos campos son producidos por una fuente fija. El tamaño de la fuente está relacionado con el volumen de trabajo. El receptor se une al objeto y tiene tres antenas perpendiculares. La distancia se deduce de los voltajes inducidos en las antenas. Se necesita calibración. Son pequeños y ligeros.

Existen dos tipos diferentes de *trackers* magnéticos: los que utilizan corriente alterna (CA) y los que utilizan corriente continua (CC). Los de CC son inmunes al acero no-ferromagnético de los metales (latón, aluminio y acero inoxidable) mientras que los de CA se ven afectados. La presencia de metales ferromagnéticos afecta a los *trackers* de CC y de CA (acero dulce y ferrita) y ambos se ven afectados por el cobre. Por ello, los *trackers* de CA tienen una resolución y exactitud mejores pero no tanto alcance.

Los *trackers* ultrasónicos son dispositivos de medida sin contacto que utilizan una señal ultrasónica producida por un transmisor fijo para determinar la posición y orientación en tiempo real de un receptor móvil. Hay dos tipos: los *trackers* de fase (no se usan) y los *trackers* de TOF (*Time of Flight*). Los *trackers* de fase comparan las fases de las ondas acústicas con la fase de una onda de referencia. El error se va acumulando. Se necesita una fuente externa de vez en cuando para corregirlo. Los *trackers* de TOF usan ultrasonidos de baja frecuencia para medir la posición. Sonido producido por una fuente triangular fija (con 3 altavoces). El receptor es triangular tiene tres micrófonos y está en el objeto. A partir del tiempo de vuelo estimado se calcula la distancia de cada altavoz a todos los micrófonos. Se usa triangulación. La distancia se calcula multiplicando el “tiempo de vuelo” por la velocidad del sonido en el aire.

Los *trackers* ópticos son dispositivos de medida sin contacto que utilizan detección óptica para determinar la posición y orientación en tiempo real de un objeto. Sus principales

características son su mayor exactitud cerca del volumen de trabajo, tienen una superficie muy grande y resistencia a las obstrucciones visuales (línea de la vista), aunque hay problemas si todos los leds están ocultos a los sensores. Son sensores pesados (no es problema en los *outside-looking-in*), escalables tienen *beacons* (faros) activos (emiten luz) o pasivos (reflejan luz). Tienen una alta tasa de actualización, un gran volumen de trabajo y no se ven afectados por los metales. Pero por otro lado tienen problema de la línea de visión, luz ambiental, radiaciones infrarrojas y son muy caros.

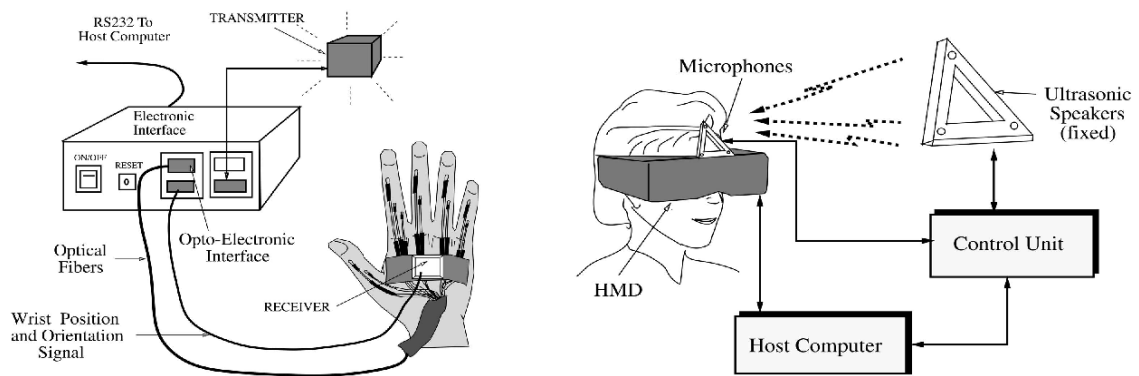


Figura 10. Ejemplos de *trackers* magnéticos (derecha) y ultrasónicos (izquierda).

Los *trackers* inerciales aprovechan propiedades físicas asociadas al movimiento para extraer aceleraciones y giros de los objetos. Los *trackers* inerciales usan acelerómetros o giroscopios. Los que usan acelerómetros son dispositivos que miden la aceleración experimentada por un objeto a partir de medir cambio en la velocidad. Estos dispositivos son hoy en día muy pequeños, de masa muy pequeña para reducir el error por el momento de inercia del propio acelerómetro. Los que usan giroscopios detectan la orientación de masas en movimiento (girando). Estos se basan en el principio de que cualquier objeto en movimiento de rotación tiene propiedades giroscópicas (conservación del momento angular). Tienen masas pesadas rotando a alta velocidad sobre unos rodamientos con baja fricción. En general, los *trackers* inerciales no necesitan ningún tipo de fuente independiente y tampoco necesitan cables y el dispositivo no está limitado en su movimiento. Además tienen un volumen de trabajo muy amplio y una baja latencia. La gran desventaja es que presentan derivas, como histéresis en acelerómetros mecánicos y tienen una mala detección de movimientos lentos.

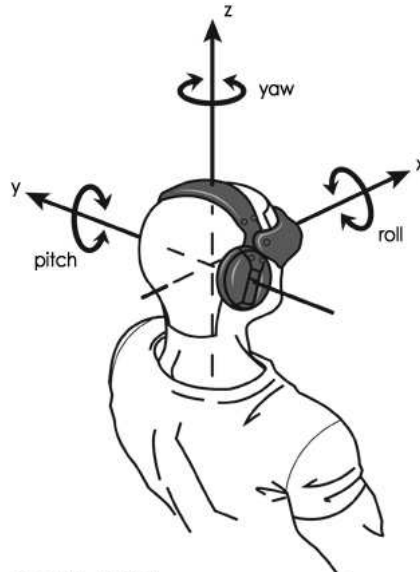


Figura 11. Ejemplo de *tracker* con sistema de referencia inercial.

Como se acaba de ver, existe una gran variedad de productos en el mercado con diferentes características, precios, funcionalidades, tamaños... Tras realizar una búsqueda y estudio de los diferentes dispositivos de posicionamiento, vamos a utilizar el RAZOR 9 de la empresa SPARKFUN.

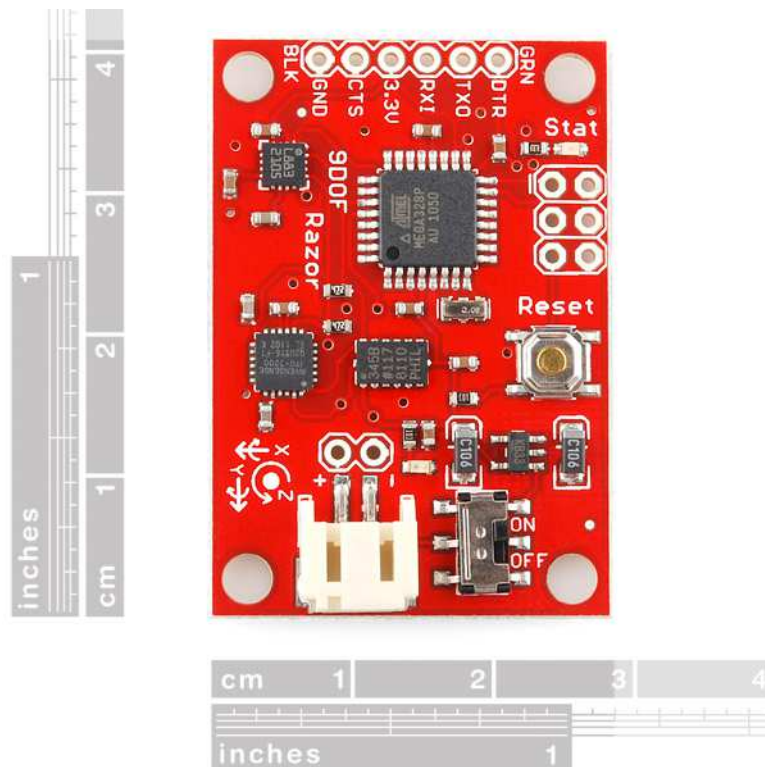


Figura 12. 9DOF Razor de la empresa SPARKFUN.

La placa 9 DOF Razor tiene IMU incorpora tres sensores - un 3200 ITG-(triple eje de giro), ADXL345 (triple eje acelerómetro), y HMC5883L (triple eje magnetómetro) - para darle nueve grados de medición inercial o ‘grados de libertad’ (*Degrees Of Freedom, DOF*). Las salidas de todos los sensores son procesadas por una placa ATmega328 y salida a través de una interfaz en serie. Esto permite al 9DOF Razor ser utilizado como un mecanismo de control muy poderoso para UAVs, vehículos autónomos y sistemas de estabilización de imagen.

El 9DOF trabaja a 3.3VDC, cualquier potencia suministrada al conector blanco JST será regulado por la presente tensión de servicio. El encabezado de salida está diseñado para acoplar la placa al puerto USB de un ordenador. O, para una solución inalámbrica, puede ser conectado al MATE Bluetooth o un Explorer XBee.

Este sensor es extremadamente pequeño con unas dimensiones de 28mm de largo y 41mm de ancho. Además incorpora un soporte por el cual se puede fijar sin problemas en cualquier base para trabajar con él.

### V.3. DESARROLLO DE LA APLICACIÓN

En este apartado vamos a describir como hemos unido entre sí todas las partes de la aplicación que hemos ido desarrollando hasta ahora. Hemos realizado un programa en MatLab que realiza todos los pasos necesarios para realizar la reproducción de una señal de audio anecoico de manera que a pesar de que el oyente realice cambios de posición y orientación, se crea una sensación acústica como si la señal de audio estuviera en una posición fija en el espacio. A continuación explicamos como funciona este sistema y nos ayudamos con un diagrama de flujo en el que se representan todos los procesos de la aplicación.

El sistema funciona de la siguiente manera. Partimos de una señal de audio anecoica que habrá que procesar para ser transmitida por la fuente acústica o emisor. Para que el efecto sea más reconocible por el usuario, es importante que se trate de una señal que no contenga muchos cambios de amplitud, y que más o menos sea constante. Un buen ejemplo sería una voz de radio hablando con un tono constante.

El siguiente paso es la inicialización, estableciendo en primer lugar el tamaño del búfer con el que trocaremos la señal entera y que será el bloque mínimo de entrada al sistema para ser procesado. La señal se trocea según el tamaño del búfer en partes mucho más pequeñas que van a ser las señales de audio de entrada del sistema. A continuación se inicializa la posición del emisor que se mantendrá fija y la posición del receptor que se

deberá actualizar en función del movimiento del oyente. El tracker es el encargado de leer la posición y orientación del receptor. Con estos datos de posición y orientación, se estima la orientación y distancia relativa entre emisor y oyente y se busca en la base de datos de HRTF la señal más cercana a la orientación del receptor y en la base de datos de RIR la señal más cercana a la posición del receptor. De esta manera obtenemos las señales de  $HRTF_{right}$ ,  $HRTF_{left}$  y  $h_{sala}$ . Con estas señales vamos a generar la  $h_{total}$  de cada canal de audio y vamos a convolucionarla con el trozo de señal de entrada de audio para crear el efecto.

Una vez terminado este proceso se reproduce el ‘trozo’ de señal de audio de salida que va a causar el efecto deseado en el usuario a través de los auriculares. Por último, se comprueba si queda algún bloque más de la señal para repetir el proceso de nuevo. A continuación se muestra el diagrama de flujo de la aplicación:

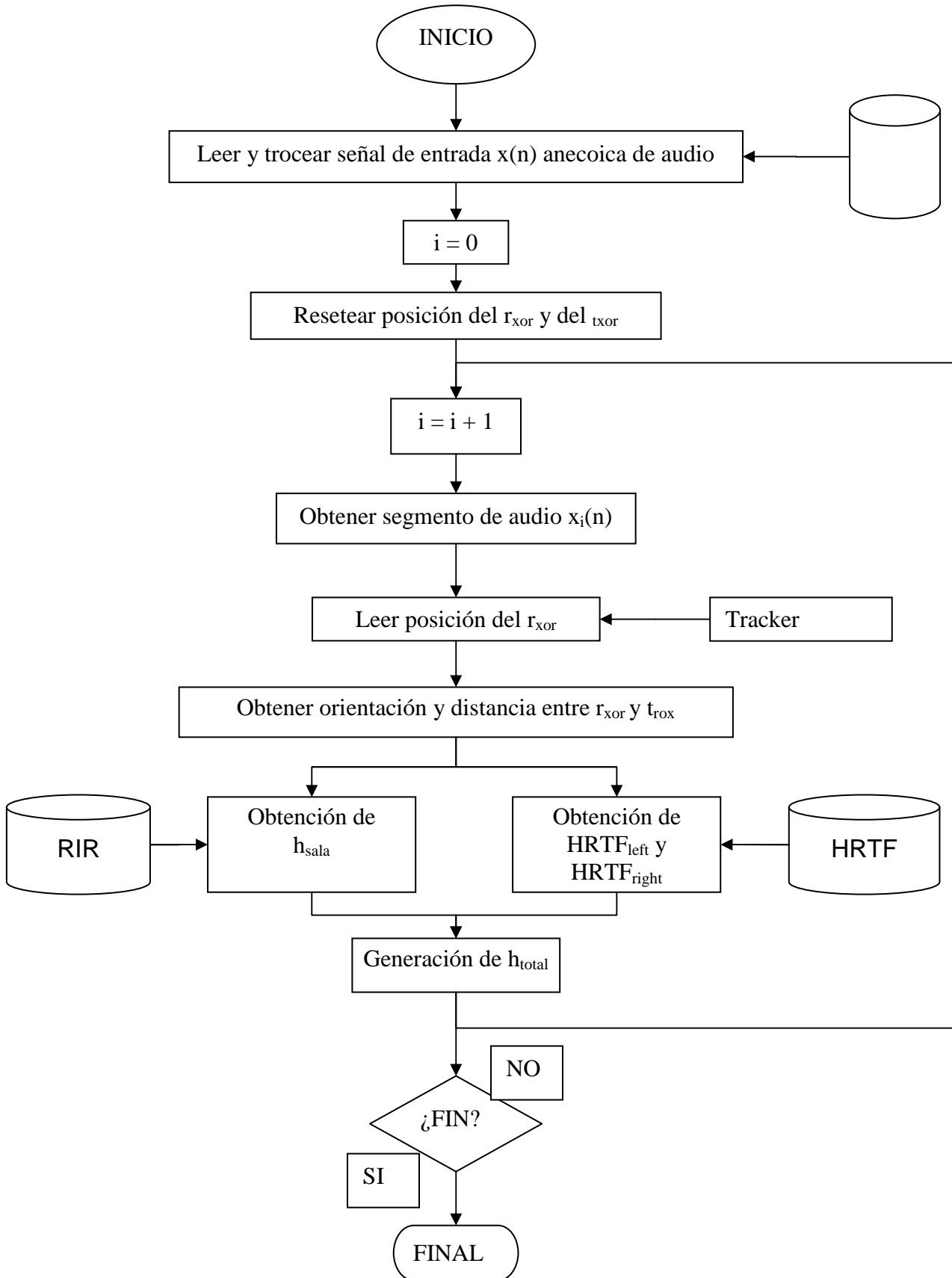


Figura 13. Diagrama de flujo de la aplicación desarrollada.



#### *V.4. APLICACIONES*

Se ha creado una aplicación capaz de crear un espacio acústico virtual, en los que la fuente o el oyente pueden realizar movimiento, creando los efectos acústicos derivados correspondientes, tal y cómo si ese movimiento se realizara en un tiempo y un espacio reales. Es decir, se trata de un sistema de realidad acústica aumentada que es capaz de ser reproducido en tiempo real.

Esto es un avance en las simulaciones acústicas y tiene múltiples aplicaciones en campos como la telefonía, videoconferencias, videojuegos, audioguías... Dentro del campo de las videoconferencias, podríamos pensar en las videoconferencias con hologramas, ya que se podría recrear una conversación casi natural, con un holograma tanto visual como acústico. De esta manera el sistema adaptaría el sonido que supone que el holograma estaría emitiendo conforme este realizara movimientos durante la videoconferencia. Sería como si la persona con la que estás hablando estuviera dentro de tu casa, situada en un punto fijo y uno pudiera moverse por la sala experimentando los efectos característicos al alejarse o girarse con respecto al holograma.

Otro de los campos que se benefician de esta aplicación es el ocio. Se puede aplicar este efecto a consolas o videojuegos de manera que el usuario puede experimentar como el sonido del entorno virtual se adapta también a sus movimientos y acciones, pudiendo mantener una fuente acústica en un punto fijo e independientemente del movimiento del jugador o desplazarla por la sala a voluntad.

También tiene aplicaciones en campos culturales y turísticos, como podría ser en audio guías de museos o atracciones. De esta manera la pintura, escultura o atracción podrían tener un sensor que detectara cuando un usuario se ha acercado y el usuario empezaría a escuchar la explicación correspondiente. Además, el sensor detectaría los movimientos del usuario, y se podría crear el efecto de que el sonido viene directamente del cuadro, cuando estemos moviendo la cabeza alrededor de él.

## **VI. CAMPAÑA DE MEDIDAS**

### *VI.1. INTRODUCCIÓN*

Para comprobar la validez de la aplicación y valorar subjetivamente su calidad, se realiza una campaña de test en la que vamos a ir validando las diferentes partes del programa para al final, juntarlo todo y poner a prueba la aplicación completa. Por un lado vamos a comprobar si las

señales de HRTF de la base de datos que hemos escogido ofrecen una buena respuesta en diferentes personas, y crean el efecto esperado de holofonía del sonido. Por otro lado queremos comprobar si la aplicación que simula la respuesta impulsiva de la sala, ofrece también buenos resultados, creando el efecto de externalización del sonido. Y, aprovechando este test, vamos a comprobar cuál es el mejor método para integrar las dos señales  $h_{sala}$  y HRTF si con el método aditivo o con el convolutivo para obtener el efecto de auralización final. Para ello generaremos las señales del test para cada uno de los métodos, convolutivo y aditivo, para ver así, qué método da mejores resultados.

Para la generación de sensación de movimiento vamos a crear un vector de movimiento, con el cual vamos a desplazar la fuente sonora. Aunque el objetivo del trabajo es en principio con una fuente fija y es el receptor el que se mueve, hemos considerado más cómo para la realización del test hacerlo a la inversa ya que el efecto es el mismo.

Las señales de audio finales tendrán una duración y un desplazamiento suficientes para que el movimiento pueda ser identificado. Si el movimiento no es muy grande, o la duración de la señal es muy corta, los sujetos bajo test serán incapaces de percibir dicho movimiento. También es importante la velocidad del movimiento ya que velocidades angulares elevadas serán difíciles de identificar con claridad por los sujetos. De este modo establecemos unos criterios a partir de los cuales vamos a crear las señales de audio para el test. La señal a escuchar será de unos 10 segundos de duración, y en el caso de haber movimiento rotatorio, éste no será menor a un cuarto de vuelta, ni mayor a una vuelta completa. De este modo aseguramos que el oyente pueda reconocer el movimiento.

El test se realiza a 10 personas de entre 24 y 28 años de las cuales 4 son mujeres y 6 son hombres. La señal de audio anecoica utilizada corresponde a una voz de mujer hablando en inglés. Esta señal se ha obtenido a partir de la página web de la empresa suiza Holophonics. Los auriculares utilizados para el test son los SONY-MDREX35LPB.



Figura 14. Durante la realización de la campaña de medidas.

## VI.2. TEST HRTF

En la primera parte del test vamos a comprobar si las señales de HRTF de la base de datos tienen una buena calidad subjetiva. Lo que se realiza en el test es utilizar estas señales de HRTF en una señal de audio anecoico. Para ello se propone simplemente crear unas señales de audio que van a tener un efecto de rotación alrededor de la cabeza gracias a un vector de movimiento que vamos a establecer previamente.

El proceso funciona de la siguiente manera. Se toma la señal de audio anecoico que va a trocearse en pedazos de tiempo muy pequeños. A cada pedazo se le atribuye una posición correspondiente dentro del vector de desplazamiento. De esta manera se calcula para cada trozo la señal de HRTF correspondiente, según la orientación en grados respecto del receptor. Se hace una aproximación ya que la base de datos de señales de HRTF es discreta y no tenemos una señal para cada ángulo. El siguiente paso es convolucionar la señal de entrada con las señales de HRTF seleccionadas. Esta convolución simula que el sonido viene de la posición dada del vector de movimiento. Gracias a la función *filter*, se puede hacer este proceso con gran facilidad ya que la esta función devuelve el estado inicial para el siguiente segmento. Estas son las condiciones iniciales del filtro utilizadas para los cálculos del siguiente segmento. El vector de movimiento se recorre entero creando una sensación de movimiento giratorio alrededor de la cabeza.

Para este primer test, se parte de una señal de unos diez segundos de duración. Se generan dos señales, la primera señal realiza un desplazamiento desde  $-90^\circ$  hasta  $90^\circ$  (desde la izquierda hasta la derecha) y la segunda señal realiza un movimiento desde  $0^\circ$  hasta  $180^\circ$  (desde delante hasta detrás por el lado derecho). El plano Z se mantiene constante para simplificar el test, y las señales siempre se encuentran a la misma distancia del oyente (1 metro).

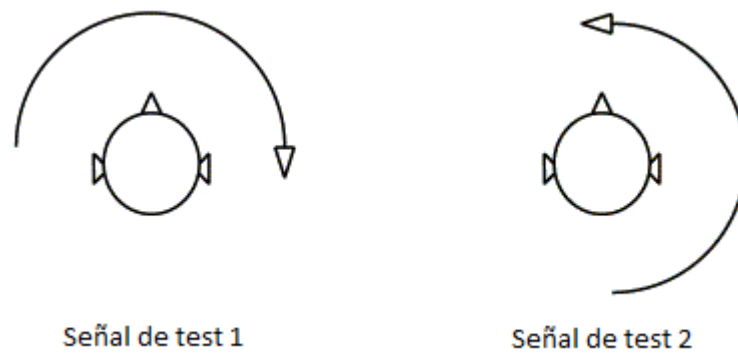


Figura 15. Esquema de la trayectoria de las señales simuladas para el test.

A partir de aquí, el sujeto tendrá que identificar el movimiento circular alrededor de la cabeza a  $0^\circ$  en el plano de elevación, que puede ser desde  $1^\circ$  hasta  $360^\circ$  en el plano de azimut. Los sujetos han de estimar el inicio y el final del movimiento, así como la dirección del mismo. Se considera una respuesta válida si coinciden la dirección del movimiento y por lo menos el 50% de la trayectoria de la señal.

En total son 20 señales de test, de las cuales 14 hacen un dibujo bastante similar a las señales simuladas en el test, considerándose como válidas. Así podemos concluir que las señales de HRTF utilizadas nos sirven, y podemos utilizarlas para los próximos procesos.

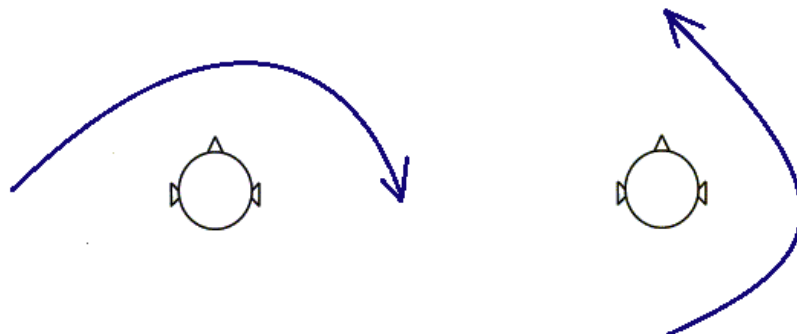


Figura 16. Resultados correctos del test.

### VI.3. TEST RIR O REVERBERACIÓN

A continuación pasamos al siguiente test realizado, que trata de comprobar la calidad subjetiva de las señales RIR de nuestra base de datos escogida, así como la integración de las señales RIR y las HRTF según el modelo aditivo o el convolutivo. Para ello, se sigue más o menos el proceso anteriormente explicado. Vamos a crear unas señales de audio para el test, a partir de señales de audio anecoicas. En este caso, la fuente sonora ya no va a realizar un movimiento alrededor de la cabeza, sino que va a estar en una posición fija. A partir de ahí, vamos a escoger las señales de HRTF y RIR correspondientes a la posición y orientación del receptor con respecto del emisor y el tipo de sala, y vamos a integrar ambas señales mediante los dos modelos de integración de la  $h_{\text{sala}}$  y las HRTF propuestos: el aditivo y el convolutivo. Vamos a crear unas señales de audio a diferentes distancias del receptor y cada señal va a ser generada por cada uno de los métodos mencionados. De esta manera, podremos comprobar que método es más fiable o se acerca más al efecto real a partir de con qué método hay mejores resultados en el test. También podremos comprobar si la aplicación de generación de las señales RIR es adecuada si los resultados del test reflejan esa espacialidad de las señales.

Las señales generadas son en el lado izquierdo a un metro y a seis metros del receptor, y lo mismo para el lado derecho. Estas cuatro señales se reproducen para ambos métodos de integración. En test consta en total ocho señales.

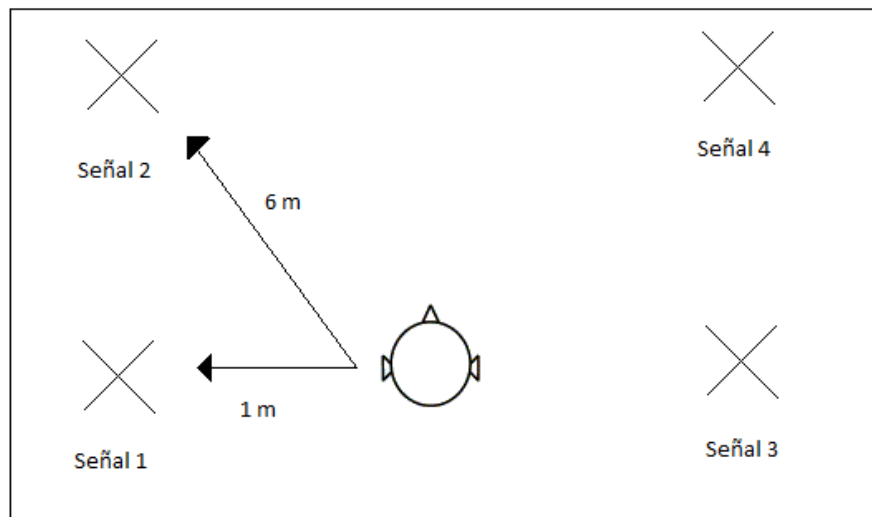


Figura 17. Esquema de las posiciones de las señales simuladas para el test.

Los sujetos bajo test tienen que identificar la posición y distancia de la fuente sonora en cada una de las señales y para cada uno de los modelos de integración. Se divide el espacio de la sala en cuadrantes representados en la plantilla de test. Se considera una respuesta válida aquellas respuestas que el dibujo esta dentro del cuadrante correspondiente a la señal de test.

En total son 80 señales de test, de las cuales 48 respuestas se consideran válidas, consiguiendo identificar la procedencia de las señales simuladas. Por eso determinamos como válidas la aplicación que genera las señales RIR.

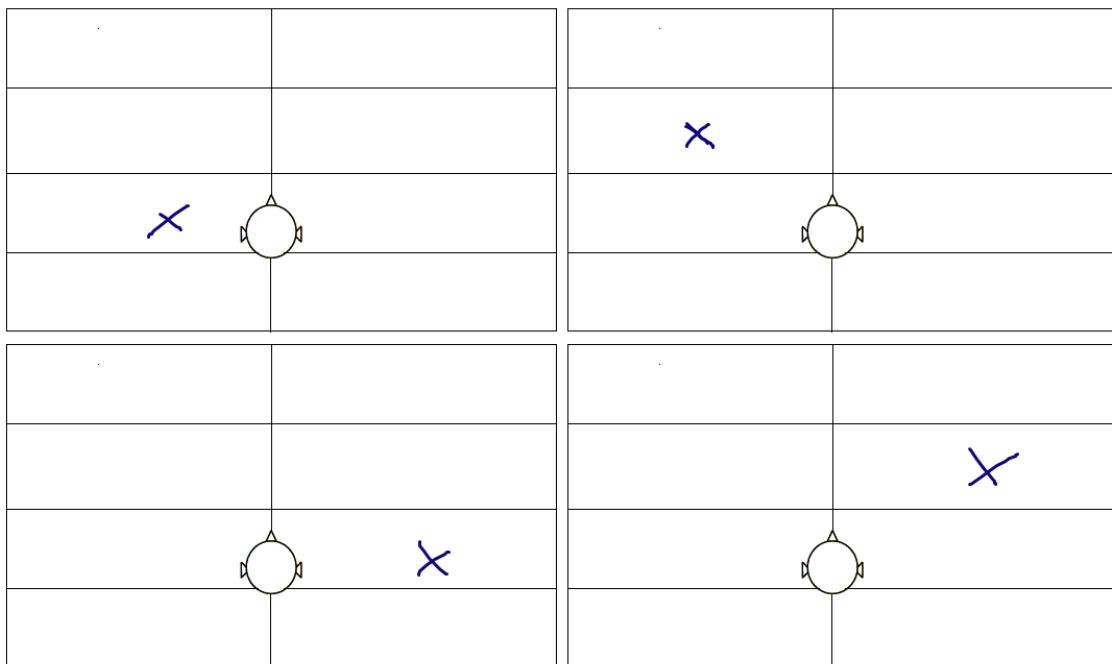


Figura 18. Resultados correctos del test.

En cuanto al método de integración, el modelo convolutivo consigue mejores resultados que el aditivo. De las ochenta señales reproducidas en el test, consideramos como válidas cincuenta y seis señales, de las cuales 27 corresponden a las generadas por el modelo convolutivo, y 21 son validas en el modelo aditivo.

#### VI.4. TEST FINAL

El último de los test, es una combinación de los dos primeros. Se trata de generar señales de audio a partir de señales de audio anecoico, un tipo de sala dado, y un vector de movimiento, para así, comprobar si realmente el efecto de movimiento dentro de una sala es bien percibido por el usuario. Para la realización de este test necesitamos añadir un vector de movimiento a la fuente sonora del test anterior. De este modo, creamos un efecto de desplazamiento de la fuente

alrededor de la cabeza, a la vez que un desplazamiento de la fuente en distancia, ya sea alejándose del oyente o acercándose. Este test va a ser generado con dos métodos, en diferido y en tiempo real.

Para el test se crean dos señales de audio con dos trayectorias diferentes. La primera señal de audio realiza un movimiento elíptico alejándose del centro de la cabeza. El sonido realiza una vuelta completa alrededor de la cabeza, desplazándose desde 1 metro de distancia hasta 6 metros de distancia. La segunda señal consta de dos movimientos: un movimiento circular de media vuelta a una distancia de un metro de la cabeza, y después un movimiento rectilíneo alejándose de la cabeza hasta los 6 metros. En la siguiente figura se ilustran las dos trayectorias.

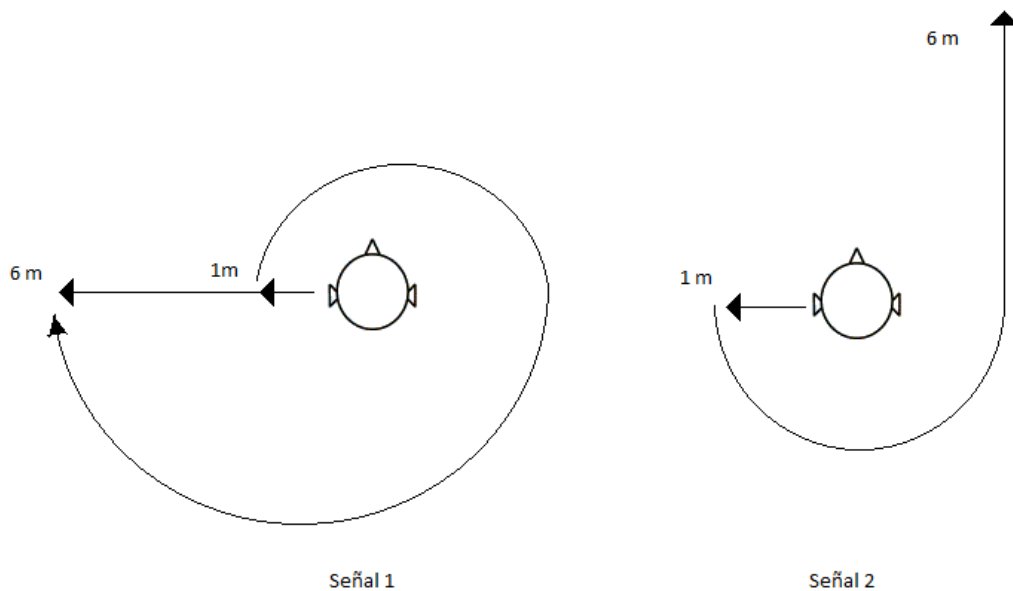


Figura 19. Esquema de la trayectoria de las señales simuladas para el test.

Para la generación de estas señales necesitamos una sala suficientemente grande. Para el test hemos simulado una sala de 10 metros de largo, 8 de ancho y 3 de alto.

Lo que queremos comprobar es como responden los oyentes ante sonidos con variaciones en las tanto en la orientación como en la distancia. El sentido de hacerlo en diferido y en tiempo real, es para comparar ambos resultados y determinar así si el procesado en tiempo real es suficientemente bueno para detectar estos movimientos o el oyente no los reconoce ni siquiera con un procesado en diferido.

Como el primer test, vamos a dar por resultados válidos los que acierten la distancia y por lo menos el 50% del recorrido del movimiento.

En total son 20 señales de test con el procesado en tiempo real, de las cuales 11 respuestas se consideran válidas, consiguiendo identificar la trayectoria de las señales simuladas. Por eso determinamos como válidas la aplicación que genera las señales RIR.

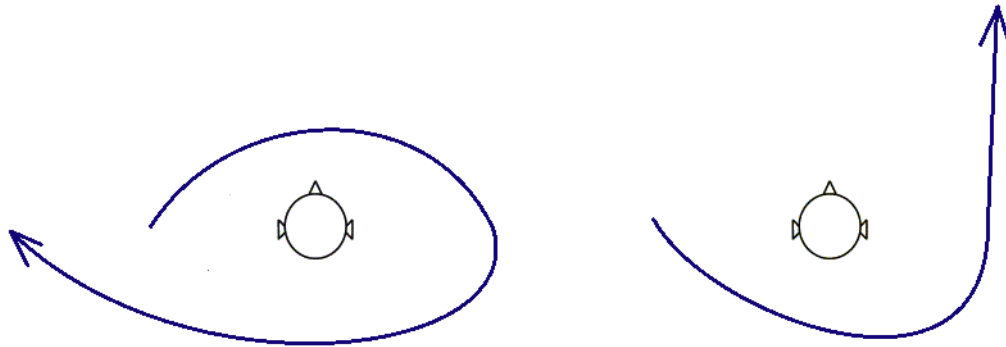


Figura 20. Resultados correctos del test.

A continuación mostramos los resultados de todos los test en una tabla.

	Nº sujetos	Nº trayectorias/ posiciones	Nº Señales	Bien	Mal
Test 1 HRTF	10	2	20	14	6
Test 2 H <sub>sala</sub> + HRTF (Método 1)	10	4	40	27	13
Test 2 H <sub>sala</sub> + HRTF (Método 2)	10	4	40	21	19
Test 3 H <sub>sala</sub> + HRTF + desplazamiento (Diferido)	10	2	20	10	10
Test 3 H <sub>sala</sub> + HRTF + desplazamiento (Tiempo real)	10	2	20	11	9

Tabla 2. Resultados de la campaña de test realizada.

## VII. CONCLUSIONES Y LÍNEAS FUTURAS

Como conclusiones generales de este trabajo podríamos decir que se ha hecho un estudio de los modelos de generación de ruido tridimensional reproducidos mediante unos auriculares, el cual lo hemos diferenciado en dos grandes efectos, el efecto de holofonía, generado mediante las HRTF y el efecto de externalización, obtenido mediante un modelo RIR o reverberante



( $h_{sala}$ ) . Estos dos efectos se han combinado de diferentes maneras, y se ha comprobado gracias a la campaña experimental que el método convolutivo de integración de las HRTF con la  $h_{sala}$  es el más adecuado, creando un efecto mas agradable para el usuario. También, se ha evaluado de manera independiente las HRTF y las RIR, y se han obtenido resultados satisfactorios, donde la sensación de inmersión acústica ha sido alta.

Finalmente como novedad de este trabajo, se ha presentado la adaptación de estos efectos al movimiento del oyente, en primera instancia a partir de un movimiento simulado y en segunda instancia mediante un posicionador o tracker. Esto permite, independizar la ubicación de la fuente acústica (dinámica o estática) del movimiento de la oyente a partir de un sonido reproducido desde unos auriculares abriendo un gran campo de aplicaciones.

Como apartado final al presente trabajo se enumeran a continuación las vías que se podrían seguir para continuar profundizando en el trabajo presentado:

- Se plantea la posibilidad de realizar un test psicoacústico más exhaustivo, realizandolo con sonidos con distinta carga espectral ya que es bien conocido que la localización de los sonidos depende de la frecuencia de los mismos. Así se obtendría una ampliación de los resultados obtenidos en este trabajo.
- Se podría realizar un estudio de los métodos de interpolación de las señales HRTF y RIR, ya que ha sido necesario hacer una discretización espacial de este tipo de señales para tener una base de datos de un tamaño manejable. De esta manera el efecto simulado sería un poco más preciso y evitaría pequeños saltos en las reproducciones.
- También sería una buena posibilidad implementar el cálculo de la RIR en tiempo real. De esta manera se podría crear el efecto en cualquier punto de la sala sin necesidad de una base de datos previa de las respuesta al impulso de la sala, y ofreciendo mucha más flexibilidad a la hora de ser aplicado en diferentes ámbitos y espacios.

## **AGRADECIMIENTOS**

Me gustaría agradecerle a mi director de Tesis Jorge Gosálbez Castillo el gran apoyo moral y académico que me ha ofrecido. También mencionar a las personas que han realizado el test por su paciencia. Además, dar las gracias a las personas que me han ayudado de una u otra forma en la realización de esta Tesina, ellos saben quienes son.

**BIBLIOGRAFÍA**

- [1] Fontana, S., Farina, A., Grenier, Y. *Binaural for popular music: Case of study*. 13th International Conference on Auditory Display, Montréal, Canada, June 26-29, 2007.
- [2] Zhong, X. L., Xie, B. S. *Progress on the research of Head-Related Transfer Function*. Audio Eng, 12: 44—46, 2004.
- [3] Recuero, M. “*Ingeniería Acústica*”. Capitulo 11. Editorial Paraninfo. 1994.
- [4] Rayleigh, Lord (1907), “*On our perception of sound direction*”, Philos. Mag. 13, 214-232
- [5] Litovsky, R.Y., Colburn, S.H., Yost, W.A., Guzman, S.J. *The precedence effect*. J.A.S.A., 106. 1633-1654. 1999.
- [6] Zurek, P.M. *The Precedence Effect and its possible role in the avoidance of interaural ambiguities*. J.A.S.A., 67. 952-964, 1980.
- [7] Freyman, R. L., Clifton, R. K., Litovsky, R. Y. *Dynamic processes in the precedence effect*. J.A.S.A., 90. 874-884, 1991.
- [8] Blauert, J. *Spatial Hearing*. The MIT Press, Cambridge: 1997. Revised Edition.
- [9] Shaw, E.A.G. *The External Ear: Handbook of Sensory Physiology V/1: Auditory System, Anatomy Physiology*. Springer-Verlag, New York: 1974.
- [10] Wightman, Frederic L. and Kistler, Doris J. *Headphone simulation of free-field listening. I: Stimulus synthesis*. J.A.S.A., 85(2): February 1989.
- [11] Kahana, Yuvi et al. *Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear*. 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction, Rovaniemi, Finland: 1999.
- [12] Brungart, D. S. and Rabinowitz, W. R., “*Auditory localization in the near-field*”, Proc. Int. Conf. Auditory Display (ICAD '96), Palo Alto, California, 1996.
- [13] Duda R. O. and Martens W. L., “*Range-dependence of the HRTF for a spherical head*”, Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, New York, Session 6.2, 1997.
- [14] Calamia, P., and Hixson, E., J. Acoust. Soc. Am. 102, (5, pt. 2), 3117 (1997).
- [15] Oppenheim, A. V., Schafer, R. W. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey: 1989.
- [16] Cheng, C. I., Wakefield, G. H. *Introduction to Head-Related Transfer Functions (invited tutorial)*. University of Michigan, Department of Electrical Engineering and Computer Science. U.S.A.
- [17] Warusfel O. *Listen HRTF Database*. Room Acoustics Team, IRCAM Institute, Paris. 2003.
- [18] Riederer K. *Repeatability analysis of head-related transfer function measurements*. 105th Convention of Audio Engineering Society, preprint paper 4846. 1998.
- [19] Hammershoi D., Moller H. *Sound transmission to and within the human ear canal*. J.A.S.A., 100. 408-427. 1996.