



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA DE  
VALPARAÍSO



**Mauricio Sebastián Lorca Cerda**

**Preparación de setup para  
experimento de localización de audio  
3D en tiempo real**

**Informe Proyecto de Título de Ingeniero Civil Electrónico**



**Escuela de Ingeniería Eléctrica  
Facultad de Ingeniería**

**Valparaíso, 4 de marzo de 2019**



# Preparación de Setup para experimento de localización de audio 3D en tiempo real

Mauricio Sebastián Lorca Cerda

Informe Final para optar al título de Ingeniero Civil Electrónico,  
aprobada por la comisión de la  
Escuela de Ingeniería Eléctrica de la  
Facultad de Ingeniería de la  
Pontificia Universidad Católica de Valparaíso  
conformada por

Sr. Sebastián Fingerhuth Massmann  
Profesor Guía

Sr. Gabriel Hermosilla Vigneau  
Segundo Revisor

Sr. Jorge Mendoza Baeza  
Director de Escuela

Valparaíso, 4 de marzo de 2019

*Dedicado a todas las personas que en algún momento creyeron en mí...*

# Agradecimientos

Ha sido un largo y difícil camino desde que todo esto empezó. Hoy puedo mirar hacia atrás y ver que todo el esfuerzo, sudor y lágrimas valieron la pena. Sin duda que lograr esto fue mucho más fácil de llevar con la ayuda de muchas personas importantes en mi vida.

Quiero empezar por agradecer a todos mis compañeros que alguna vez me ayudaron, enseñaron y apoyaron en algún momento académico, pero por sobre todo a los que me regalaron parte de su tiempo, forjando una linda amistad que perdurara por siempre. A mis abuelos, tíos y primos por entregarme siempre todo lo que he necesitado para realizarme como persona. A mis padres Mauricio y Jessica por su apoyo incondicional. A mi polola Claudia por traerme luz y apoyo en esta última parte de esta etapa.

A Pedro Pablo “El Tucu” Hernández por darme la oportunidad de presenciar el gol más importante en la historia del club O’Higgins de Rancagua. A Pablo Ignacio Calandria por su entrega y garra incondicional. A Eduardo Berizzo por darnos un título y la felicidad más grande que pueda guardar un Rancaguino Celeste. Y a Don Marcelo Bielsa por darnos catedra de vida y humildad.

Finalmente nada de esto sería posible sin la ayuda y consejos de mi profesores guía el Dr. Sebastián Fingerhuth y el Dr. Gabriel Hermosilla Vigneau. Agradezco cada momento que se dieron para poder ayudarme y guiarme durante este proceso. Por otro lado agradecer también al FONDECYT #11150388 por darme la oportunidad de trabajar y financiar este hermoso proyecto que, sin duda, disfruté mucho.

*Valparaíso, 3 de enero de 2019*

Mauricio Lorca

“Hay un día en que el soñador y el sueño se encuentran, y ese día es hoy”

*26 de diciembre de 2013*

Eduardo Berizzo

# Resumen

La primera parte de este informe tiene la finalidad de exponer conceptos fundamentales para poder entender lo que es el sonido 3D y los conceptos asociados a éste. Se presentarán las herramientas que usa el ser humano para localizar fuentes sonoras a través de la audición, analizando nuestro sistema de recepción de sonido y sus limitaciones. Además, se explicarán cómo afectan para la simulación de audio 3D, los aspectos físicos de cada oyente. En base a los conceptos que se plantean, se definió una problemática y se repasaron distintos experimentos de audio 3D, que servirán como apoyo para realizar una solución a la problemática planteada.

Una vez planteado los conceptos básicos, se hablará acerca de las herramientas específicas que serán fundamentales para poder llevar a cabo el proyecto. Se estudiaron formatos de audio directivos que se podrán ocupar para almacenar las mediciones realizadas en el Laboratorio de Acústica y Vibraciones (LAV) de la Escuela de Ingeniería Eléctrica (EIE), sobre las funciones de transferencia asociadas a la cabeza. Además, se explicaron algunas herramientas de auralización asociadas con Matlab. Finalmente se investigó sobre un sensor que permita hacer un seguimiento de los movimientos de la cabeza.

Posterior a esto, se explicará cómo montar una escena de simulación de audio tridimensional con el software *Virtual Acoustic (VA)*. Se mostrará cómo configurar VA para que el servidor que procesará los datos se conecte de forma satisfactoria con la interfaz de trabajo, que en este caso será MATLAB. Además, se exponen las etapas importantes a la hora de plantear una simulación de audio tridimensional en VA.

Se realizaron pruebas para comprobar que las simulaciones de audio se hayan realizado de forma correcta, tanto para fuentes sonoras fijas en el espacio como para fuentes sonoras en movimiento constante alrededor del oyente. Por último, se muestran los resultados de una medición de “*Head-related transfer function*” (*HRTF*) realizada por el laboratorio de acústica y vibraciones de la Escuela de Ingeniería Eléctrica de la PUCV, y finalmente se agregó un seguidor del movimiento de cabeza (Headtracking).

Palabras claves: sonido 3D, localización, simulación, Virtual Acoustic, Headtracking.

# Abstract

The first part of this report aims to present fundamental concepts to understand what 3D sound is and the concepts associated with it. We present the tools that the human being uses to locate sound sources through hearing, analyzing our sound reception system and its limitations. In addition, we will explain how to make 3D audio simulation, the physical aspects of each ear. Based on the concepts raised, a problem was defined and the different 3D audio experiments were reviewed, which serve as support to make a solution to the problem posed.

Once the basic concepts are discussed, the specific tools that will be fundamental to carry out the project will be discussed. Executive audio formats were studied that can be used to store the measurements made in the Laboratory of Acoustics and Vibrations (LAV) of the School of Electrical Engineering (EIE), on the functions of transfer associated with the head. In addition, some auralization tools associated with Matlab were explained. Finally, we investigated a sensor that allows tracking the movements of the head.

Then, we explain how to set up a three-dimensional audio simulation scene with the Virtual Acoustic (VA) software. It shows how to configure VA so that the server that processes the data connects satisfactorily with the work interface, which in this case will be MATLAB. In addition, the important stages are exposed when proposing a three-dimensional audio simulation in VA.

Tests were carried out to verify that the audio simulations were performed correctly, both for fixed sound sources in space and for sound sources in constant movement around the listener. Finally, the results of a measurement of "Head-related transfer function" (HRTF) performed by the acoustics and vibrations laboratory of the Electrical Engineering School of the PUCV, and finally a follower of the head movement (Headtracking).

Keywords: 3D sound, location, simulation, Virtual Acoustic, Headtracking.

# Índice general

Resumen .....	5
Abstract .....	6
Introducción.....	1
1 Marco teórico .....	4
1.1 Definición de la problemática .....	4
1.2 Contextualización de trabajo .....	4
1.3 Sistema auditivo periférico .....	5
1.4 Localización.....	5
1.4.1 Movimiento de cabeza .....	9
1.4.2 Head-related transfer function (HRTF) .....	9
1.4.3 Sistema estereofónico .....	11
1.4.4 Reconstrucción binaural .....	12
1.5 Solución propuesta .....	15
1.6 Objetivo general .....	17
1.7 Objetivo específicos .....	17
2 Metodología.....	18
2.1 Virtual Acoustic .....	18
2.1.1 Directividades (HRTF y HRIR).....	19
2.1.2 Representación de audio en VA .....	19
2.1.3 Modo de auralización .....	20
2.1.4 Reproducción de audio en VA.....	20
2.2 Formatos .....	22
2.2.1 OpenDaff (daff) .....	22
2.2.2 SOFA .....	24
2.2.3 Especificaciones generales y convenciones .....	25
2.3 ITA Toolbox para Matlab.....	27
3 Simulación de audio 3D .....	29
3.1 Virtual Acoustic – OpenDAFF .....	29

---

3.2 Grabación de los resultados de la simulación .....	33
3.3 Análisis de resultados de la simulación .....	34
3.4 Simulación de audio 3D: SOFA.....	40
<b>4 Mediciones de HRTF y HpTF .....</b>	<b>41</b>
4.1 Mediciones de HRTF .....	41
4.1.1 HRTF: Medición cada 20° en acimut .....	42
4.1.2 ITA HRTF: Medición cada 10° en acimut .....	45
4.2 Medición de función de transferencia de audífonos .....	48
<b>5 Implementación de Head-Tracking .....</b>	<b>52</b>
5.1 Mtw Awinda – Xsens .....	52
5.2 Estados de funcionamiento del sensor MTw .....	54
5.3 Funcionamiento.....	55
5.4 Implementación.....	58
5.5 Resultados .....	62
<b>Conclusiones .....</b>	<b>64</b>
<b>Bibliografía .....</b>	<b>66</b>
<b>A Códigos de Matlab.....</b>	<b>69</b>
A.1 Proceso para cargar y visualizar datos de un archivo DAFF y SOFA .....	69
A.2 Simulación de audio tridimensional de fuente sonora fija en el espacio. ....	72
A.3 Simulación de audio tridimensional de fuente sonora en movimiento.....	73
A.4 Simulación de audio tridimensional de dos fuentes sonoras con posición fija. ....	74
A.5 Simulación de audio 3D con formato SOFA. ....	75
A.6 Proceso de asignación de coordenadas a un archivo de la clase “itaHRTF” .....	76

# Introducción

En el ser humano, la visión aporta gran parte de la información espacial, pero ésta es limitada a la dirección frontal, sin embargo, es la audición la que aporta la capacidad de “observar” el entorno. Ésta es capaz de detectar, identificar y localizar las emisiones de sonido en cualquier dirección alrededor del oyente. Además, la audición permite estimar las propiedades acústicas, como el tamaño de una habitación o saber, por ejemplo, si hay una pared cercana. Se puede afirmar entonces que percibir sonidos en tres dimensiones es una ayuda potente de supervivencia y ubicación espacial como también en el sentido del equilibrio. Una característica fundamental que le permite al ser humano determinar la localización de una fuente sonora en el espacio, es la capacidad de audición binaural. Esto es, en términos simples, la capacidad de percibir el sonido por dos oídos.

Para poder percibir un evento sonoro, se presupone la existencia de una cadena sencilla de efectos. Una fuente sonora genera vibraciones de pequeña amplitud en el aire que la rodea y debido a la compresibilidad y a la masa del aire, éstas se propagan y llegan al oído del auditor [1]. Una onda sonora, antes de alcanzar los tímpanos de un oyente, sufre una serie de transformaciones al interactuar con el torso, los hombros y especialmente con la cabeza y los pabellones auriculares, es decir, los sonidos son modificados según la geometría corporal y auditiva de manera diferente según la dirección de entrada. Para poder estudiar esta serie de transformaciones, se deben hacer mediciones con un tipo de señal que permita registrar y entender el comportamiento del par fuente-receptor. Para esto se utiliza la respuesta impulso, que se puede definir como la respuesta en el dominio de tiempo de sistema que se desea analizar bajo un impulso de corta duración. El resultado de la respuesta impulso permitirá apreciar de forma gráfica la información tanto en amplitud como en el tiempo del sistema, el cual tendrá ciertas características dependiendo de la posición de donde se encuentre, en este caso, el par fuente/receptor.

A partir de la respuesta de impulso se obtiene gran cantidad de información acerca de un sistema acústico, incluyendo los tiempos de llegada y contenido en frecuencia tanto del sonido directo como de las reflexiones. También se puede obtener información del tiempo de reverberación, la relación señal-ruido e información sobre la inteligibilidad y la respuesta en frecuencia general.

Se puede concluir que las respuestas impulsivas relacionadas con la cabeza (Head Related Impulse Response, HRIR en adelante) contienen las claves fundamentales que un ser humano utiliza para localizar una fuente sonora ubicada en el espacio que lo circunda [2] [3]. Estas son diferentes para cada oído y varían sistemáticamente con la ubicación de la fuente sonora en el espacio. Por esto, las HRIR son medidas entre posiciones diferentes de una fuente sonora, con micrófonos miniatura colocados en la proximidad de los tímpanos. Para definir la ubicación de la fuente se utiliza un sistema de coordenadas esféricas. El origen es la intersección de tres planos: el plano horizontal o acimutal, que contiene una línea imaginaria que une ambos oídos; el plano medio, que divide la cabeza simétricamente; y el plano frontal, perpendicular a los dos anteriores. Cada posición está definida por los ángulos de acimut, elevación y radio.

Por otro lado, el concepto de auralización indica que, si mediante algún medio de reproducción de sonidos se aplican a los tímpanos de un oyente las señales biológicamente correctas, será posible estimular en él la misma sensación de presencia o inmersión en el entorno modelado, es decir, una representación auditiva del espacio. Este concepto es parte del principio que sustenta la simulación 3D o simulación acústica y fue enunciado por Kleiner, Dalenbäck y Svensson, en 1993, introduciendo por primera vez el concepto de auralización [2].

En base a esto, el propósito de la simulación acústica es recrear la experiencia auditiva que una persona experimentaría en un entorno real, provocando en ella la sensación de presencia en dicho entorno, obteniendo un ambiente lo más cercano a la realidad. Los sonidos son creados por los altavoces (o auriculares), pero desde percepción del oyente, los sonidos provienen de puntos arbitrarios en el espacio. Para esto, tres subsistemas deben ser modelados: la fuente sonora, el recinto y el oyente. La fuente sonora se especifica por sus características de directividad y respuesta en frecuencia; el recinto, por la respuesta impulsiva entre una fuente sonora y un receptor; y el oyente, por las respuestas impulsivas asociadas a la cabeza (HRIR). Las transformadas de Fourier de las HRIR se conocen como Funciones de Transferencia de Cabeza (HRTF) [3].

En el mundo del audio, existen variadas técnicas para obtener un sonido 3D o señales binaurales, es decir, obtener dos canales independientes que representan el frente sonoro que llega a ambos canales auditivos, dependiendo de donde se encuentre la fuente sonora. Este trabajo se acotará a obtener o generar señales binaurales a partir de un sonido monoaural, es decir, un sonido que está definido por un solo canal. Para esto se utiliza la convolución, que básicamente se define como una operación matemática que combina dos señales para producir una tercera señal, utilizando la transformada rápida de Fourier (FFT). En el campo de las señales digitales es considerablemente importante, ya que permite obtener la señal de salida de un sistema a partir de la señal de entrada y la respuesta al impulso. Es decir, se puede predecir la salida conociendo la entrada y la respuesta al impulso. En pocas palabras, el software de convolución carga y analiza una respuesta impulso, luego, aplica la convolución con una señal que se desea transformar. Teniendo como resultado una nueva señal que da la impresión de que actúa físicamente con el espacio muestreado (donde se realizó la medición de respuesta impulso), dando como resultando simulaciones realistas.

La localización precisa de las fuentes de sonido depende de la conjunción de una serie de señales que se deben a la naturaleza del sonido, las características antropométricas y la audición del oyente, el movimiento voluntario o involuntario de la fuente o el oyente, y el entorno físico en el que se sumerge el oyente. Si bien algunas de estas señales son de valor para la localización exacta, otros, tales como ecos de habitaciones y reverberación, son a menudo perjudiciales [4].

Los oyentes utilizan instintivamente el movimiento de la cabeza para desambiguar y corregir el sonido en el espacio. Si se elimina esta habilidad, la capacidad para ubicar los sonidos en el espacio disminuye, particularmente con respecto a la elevación y la parte delantera/trasera, sobre esto se entrara en detalles más adelante. Cuando un oyente gira su cabeza 45 grados hacia un lado, se debe ser capaz de reflejar eso en su entorno auditivo, o el sonido sonará falso. Es por eso que es relevante estudiar la opción de poder incluir un seguidor del movimiento de la cabeza dentro del sistema de auralización o simulación de audio 3D [5].

Los dispositivos de seguimiento se comunican con la unidad de procesamiento del sistema, en este caso el auralizador, diciéndole la orientación del punto de vista del usuario. En los sistemas que permiten a un usuario moverse dentro de un espacio físico, los rastreadores detectan dónde está el usuario, la dirección en la que se mueve y su velocidad.

Hay varios tipos diferentes de sistemas de seguimiento, algunos de estos altamente utilizados en los sistemas de realidad virtual, pero todos ellos tienen algunas cosas en común. Pueden detectar seis grados de libertad: son la posición del objeto dentro de las coordenadas x, y, z de un espacio y la orientación del objeto en 3 ejes (roll, pitch, yaw).

# 1 Marco teórico

Este capítulo tiene el fin de poder contextualizar algunos conceptos que serán claves para poder entender el fundamento de la simulación de audio 3D. Se partirá planteando una problemática, seguido por una presentación de antecedentes generales y finalmente se dará una solución propuesta para la problemática planteada.

## 1.1 Definición de la problemática

En la simulación de sonido 3D se utilizan mediciones de HRTF. Estas pueden ser de maniqués de pruebas que representan, por lo general, a una persona en particular o algún promedio de facciones físicas de un grupo de personas. El no utilizar HRTF individualizadas para un oyente provocará problemas como la falta de precisión a la hora de localizar fuentes sonoras, debido a las geometrías anatómicas distintas que tiene cada persona. Debido a esto, se deben realizar mediciones de HRTF individualizadas, para poder hacer una comparación experimental entre HRTF individualizadas, medidas por el laboratorio de acústica y vibraciones de la EIE, y mediciones de HRTF genéricas publicadas por distintos laboratorios alrededor del mundo. Esto, a través de una plataforma de simulación de audio 3D.

## 1.2 Contextualización de trabajo

Como se explicó en la introducción, un sonido generado en el espacio, crea una onda de sonido que se propaga a los oídos del oyente. La interferencia con el sistema auditivo hará que se perciba una espacialidad en tres dimensiones [6].

Con lo que respecta a la sensación del oyente frente a una experiencia subjetiva de un suceso sonoro, destacan los siguientes factores:

- La naturaleza de la fuente sonora.
- La experiencia del oyente en la localización del sonido.
- El comportamiento del espacio acústico y la propagación en general.
- La influencia del torso, cabeza y oreja (características individuales de la persona).

Para entender un poco más cómo funciona el audio 3D, se debe comenzar por examinar cómo los humanos pueden localizar los sonidos utilizando sólo dos orejas.

### 1.3 Sistema auditivo periférico

El gran responsable de los procesos fisiológicos que conlleva la audición, es el sistema auditivo periférico. Este sistema está compuesto por el oído externo, el oído medio y el oído interno (Figura 1-1). Cumple funciones relevantes en la percepción del sonido, principalmente en la transformación de las vibraciones de presión sonora, captadas por el oído y que posteriormente llegan al tímpano, en señales neurológicas que son enviadas al cerebro a través de los nervios auditivos.

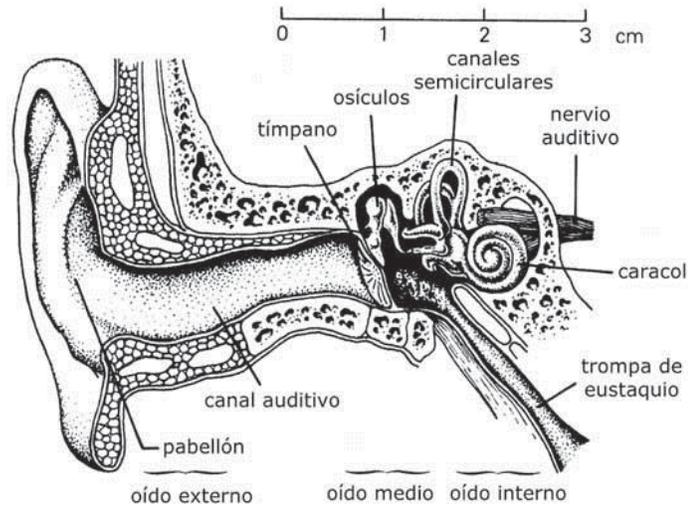


Figura 1-1: Esquema sistema auditivo [7].

**Oído externo:** en esta sección se canaliza la energía acústica. Está compuesto por el pabellón, que concentra las ondas sonoras en el conducto, y el conducto auditivo externo que desemboca en el tímpano.

**Oído Medio:** en su interior se encuentran tres pequeños huesos: el martillo, el yunque y el estribo, los cuales funcionan como amplificadores de las vibraciones que se han recibido en el oído externo. Aquí se transforma la energía acústica, canalizada por el oído externo, en energía mecánica transmitiéndola y amplificándola hasta el oído interno.

**Oído Interno:** en esta sección se realiza la definitiva transformación de la energía mecánica en impulsos eléctricos. La cóclea, también llamada caracol, es la principal responsable de esta tarea: transformar las vibraciones recibidas en las secciones anteriores a impulsos eléctricos, para que finalmente entren al cerebro para su interpretación como sonidos [7].

### 1.4 Localización

Antes de simular la escucha espacial humana es indispensable conocer qué métodos y técnicas utiliza el sistema auditivo humano para determinar la localización de fuentes sonoras en el espacio. Además, el cerebro debe aprender y calibrar estas señales utilizando los datos de precisión espacial de los otros sistemas sensoriales [8].

Existen dos sistemas fundamentales para determinar la localización de una fuente sonora en el plano acimutal (plano horizontal en figura 1-2): la diferencia de tiempo interaural y la diferencia de intensidad interaural. A estos sistemas de localización se les conoce como teoría dúplex [9]. En términos simples, cuando un sonido está más cerca de la izquierda, el oído izquierdo lo escucha antes de que el oído derecho lo escuche además de sonar más fuerte.

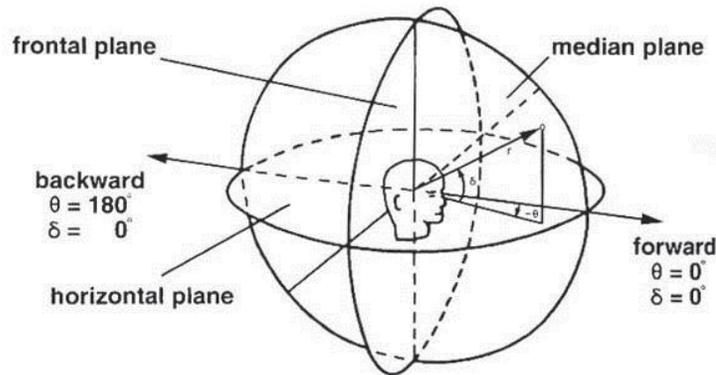


Figura 1-2: Sistema de referencia [10].

**Diferencias de tiempo interaural** (Interaural Time Differences: ITD): toma los distintos tiempos de llegada del frente de sonido, para cada uno de los oídos. Esto es posible gracias a la distancia de los dos oídos, de aproximadamente 20 centímetros. Sin embargo hay posiciones donde estas diferencias de tiempo son mínimas, provocando confusiones (Ver figura 1-3).

**Diferencias de intensidad interaural** (Interaural Level Differences: ILD): toma las distintas intensidades captadas por ambos oídos. Esto es posible gracias al umbral que provoca la propia cabeza que hace de panel acústico absorbente (Ver figura 1-3).

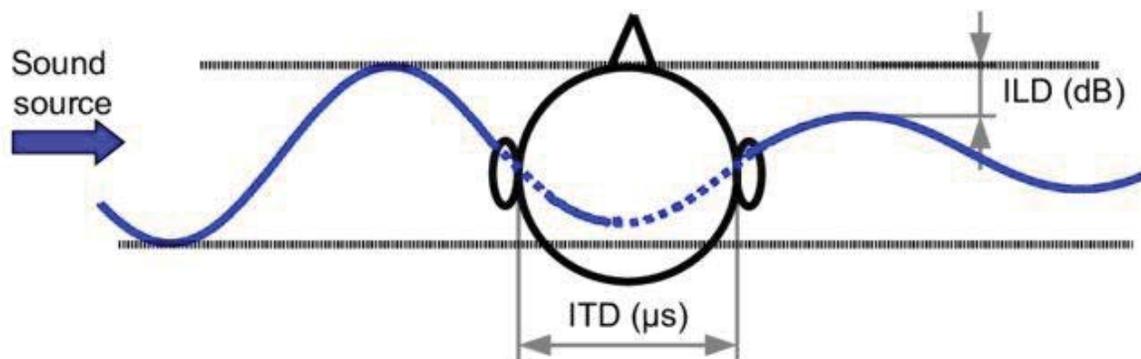


Figura 1-3. Ilustración de la diferencia de tiempo interaural (ITD) y la diferencia de nivel interaural (ILD) [11].

La localización frontal y trasera es significativamente más difícil que la localización lateral. No se puede confiar solo en las diferencias de tiempo y/o de nivel, ya que las diferencias pueden ser nulas para un sonido que se produce delante o detrás del oyente.

En la figura 1-4, se puede ver cómo los sonidos en las ubicaciones A y B, no se pueden distinguir las diferentes posiciones, ya que están a la misma distancia de ambos oídos, lo que da diferencias idénticas de nivel y tiempo.

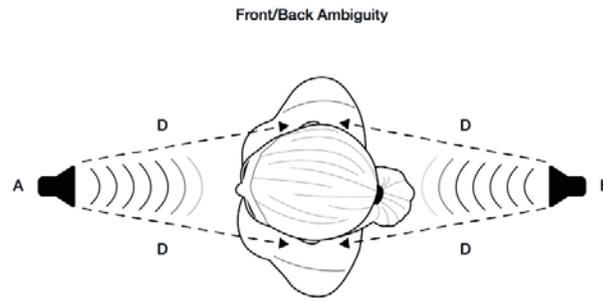


Figura 1-4. Confusión front/back [5].

De la descripción de los sistemas interaurales de tiempo y nivel, se puede observar que ambos mecanismos se complementan, transformándose en mecanismos de localización binaurales. En las bajas frecuencias predominarán las diferencias interaurales de tiempo, considerando que a partir de 1500 [Hz] la longitud de onda es comparable al diámetro de la cabeza, lo que causa que ITD empiece a ser confusa. Mientras que en altas frecuencias lo harán las diferencias interaurales de nivel, a frecuencias mayores de 1500 [Hz], la cabeza empieza a hacer sombra en el oído más alejado de la fuente de sonido, por lo que menos energía llega a ese oído, debido a esto, las ILD son esenciales para localizar sonidos con frecuencias altas (Ver figura 1-5).

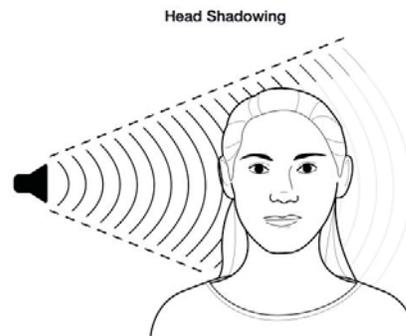


Figura 1-5. Representación del efecto sombra [5].

La localización del sonido está comprendida también en la elevación y la distancia. Existe una zona donde las ITD e ILD no especifican una única posición espacial; es más, hay una cantidad muy alta de posiciones a igual distancia del oyente que tienen las mismas ITD e ILD. Esa serie de puntos infinitos es llamado “cono de confusión”. En la figura 1-6 se puede apreciar cómo (A y

B) y (C y D) tienen las mismas ITD e ILD, ya que llegan al mismo tiempo y con la misma intensidad, haciendo imposible al receptor la localización precisa de dichas fuentes.

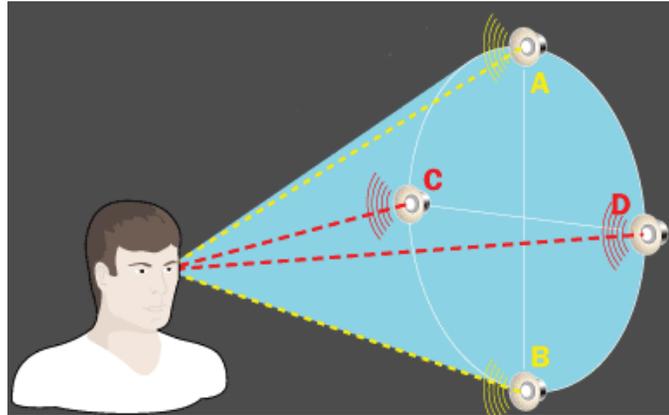


Figura 1-6: Cono de confusión [12].

### Elevación

La determinación de la elevación es netamente monoaural, es decir que se puede percibir la elevación sólo con un oído, debido a sus características de resonador acústico. Por otra parte la discriminación del ángulo lateral, como se mencionó anteriormente, es binaural.

Las cavidades del oído amplifican algunas frecuencias, y su geometría provoca efectos de interferencia que atenúan otras. Es más, su respuesta frecuencial depende de la dirección (Ver figura 1-7).

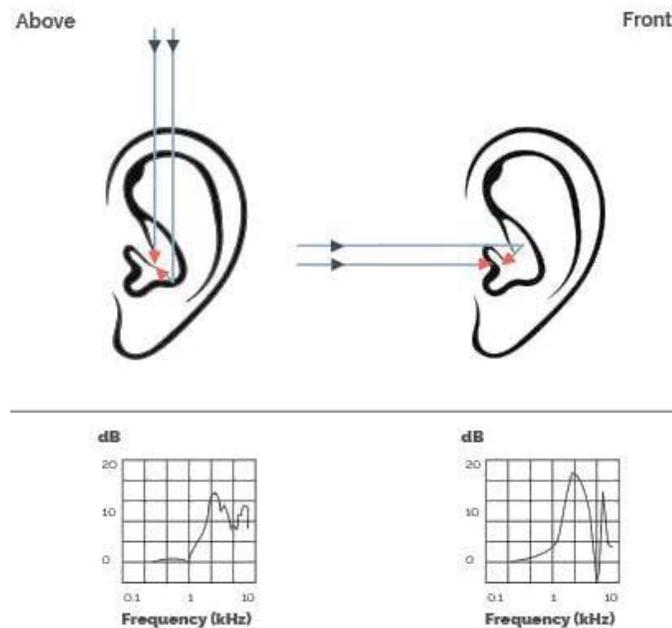


Figura 1-7: Ejemplo de respuesta en frecuencia para fuentes sonoras a diferente elevación [13].

### 1.4.1 Movimiento de cabeza

Existen ocasiones donde las formas de localización explicadas anteriormente, pueden no ser suficientes para localizar un sonido con precisión, por lo que a menudo se depende del movimiento de la cabeza. La rotación de la cabeza provoca diferencias interaurales fundamentales para determinar si una fuente de sonido se encuentra en frente o detrás.

En la figura 1-8, los sonidos en A y B no se distinguen entre sí en función de las diferencias de nivel o de tiempo, ya que son idénticos. Al girar la cabeza ligeramente, el oyente altera las diferencias de tiempo y nivel entre los oídos, ayudando a desambiguar la ubicación del sonido. D1 está más cerca que D2, lo que es una señal para el oyente de que el sonido está ahora más cerca de la izquierda y, por lo tanto, detrás de ella.

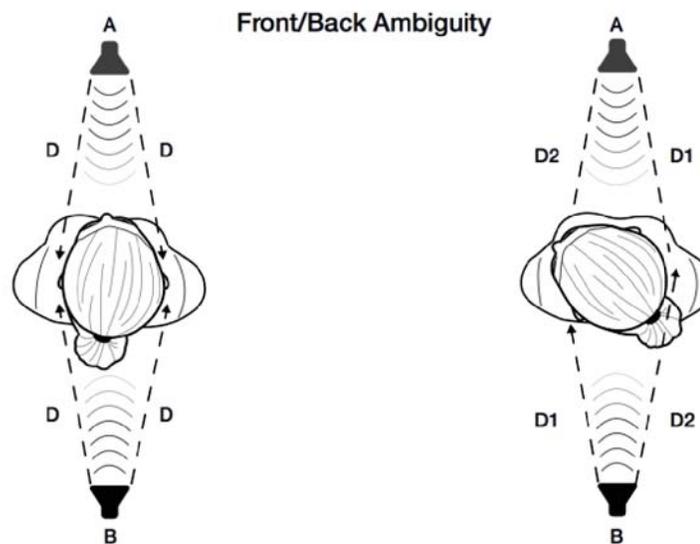


Figura 1-8: Movimiento de cabeza [5].

### 1.4.2 Head-related transfer function (HRTF)

Las funciones de transferencia relacionadas a la cabeza (HRTF) son muy usadas en el desarrollo de los sistemas de sonido tridimensional. Una HRTF se le puede considerar como un modelo que toma en consideración todas las pequeñas difracciones, resonancias y retardos que se producen por la presencia y la forma de la oreja del oyente. A las HRTF de los oídos derecho e izquierdo se las considera como dos filtros separados que reciben dos señales distintas dependientes de la posición de la fuente de sonido y presentan un retardo de tiempo entre los dos oídos [14], es decir la HRTF depende de la localización en acimut, elevación y distancia ( $\theta$ ,  $\phi$ ,  $r$ ).

Formalmente la HRTF se define como la respuesta de frecuencia del oído izquierdo o derecho de un individuo, medida desde un punto específico en el campo libre a un punto específico en el canal auditivo [15]. Un método para obtener la función de transferencia de la cabeza es medir la respuesta del impulso relacionado con la cabeza (Head Related Impulse Response o HRIR) justo

en el tímpano. Se mide, idealmente en una cámara anecoica para reducir al mínimo la influencia de las reflexiones y reverberaciones, en incrementos pequeños que pueden ir desde aproximadamente 5 a 10 grados tanto en el plano horizontal como el plano vertical, usando en algunos casos una fuente giratoria para poder manejar la variación en acimut y alguna especie de estructura que permita ir variando la posición en altura de los parlantes. Un ejemplo de este método se presenta en la figura 1-9.

Para definir la ubicación de la fuente se utiliza un sistema de coordenadas esféricas. El origen es la intersección de tres planos: el plano horizontal o acimutal, que contiene una línea imaginaria que une ambos oídos; el plano medio, que divide la cabeza simétricamente; y el plano frontal, perpendicular a los dos anteriores. Cada posición está definida por los ángulos de acimut y elevación (Ver figura 1-2).

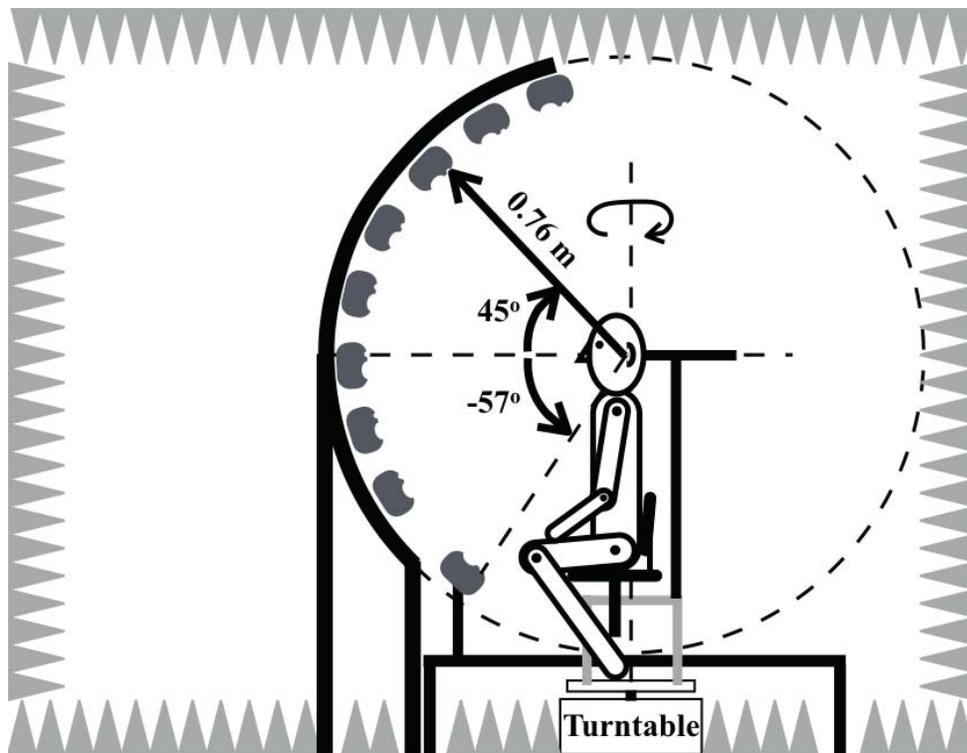


Figura 1-9: Diagrama de medición de HRTF en una cámara anecoica [16].

Una HRIR típico se puede descomponer en una serie de eventos de sonido, tal como se muestra en la figura 1-10. Primero hay un retardo de tiempo inicial debido a la distancia de la fuente con respecto a los oídos. Entonces, un pulso directo llega, cuya amplitud depende de la distancia de la fuente y el efecto sombra, seguido de una combinación de “crestas”, causadas por las reflexiones y difracciones debido a cavidades del pabellón auricular. El resto de la señal contiene reflexiones de hombro, torso y dispositivos de medición tales como el plato giratorio. Técnicamente, el pulso directo no puede ser parte de la respuesta de la pinna, pero la primera respuesta, que tiene una duración de aproximadamente 0,2 [ms] desde la llegada del impulso

directo, se conoce como la respuesta pinna [17]. La diferencia de tiempo interaural (ITD) y la diferencia de nivel interaural (ILD) se pueden evaluar con la respuesta impulsiva.

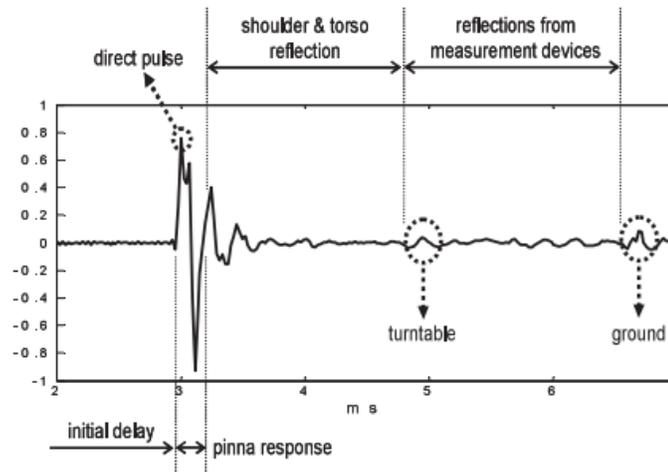


Figura 1-10: Descomposición estructural de una HRIR [17].

### 1.4.3 Sistema estereofónico

El término “estereofonía” hace referencia a cualquier sistema sonoro que maneja sonidos tridimensionales. Es necesario tener sonido estereofónico para obtener una perspectiva aural diferente dependiendo de la localización de los instrumentos sonoros. Sin embargo, hay técnicas estéreo cuya reproducción se basa en altavoces que reflejan con distinta precisión las referencias espaciales existentes en el sonido original.

#### Dos canales

En sistemas de sonido estereofónicos que tienen dos canales independientes de señales de audio, las ondas emitidas por cada parlante tienen un nivel específico y una determinada variación de fase, que varía entre los dos canales de audio, lo que da la sensación de una fuente de sonido imaginaria, es decir, que esta nueva fuente tenga posición distinta que la ubicación de la fuente de sonido monofónico.

Esto quiere decir que una señal monofónica que está difundida entre los dos canales no tiene información necesaria de la fase, mientras que una señal estereofónica tiene una diferencia entre los dos canales que simula una diferente posición de la fuente de sonido.

La configuración óptima para un sistema estereofónico es un triángulo equilátero entre los altavoces y el oyente, como muestra la figura 1-11.

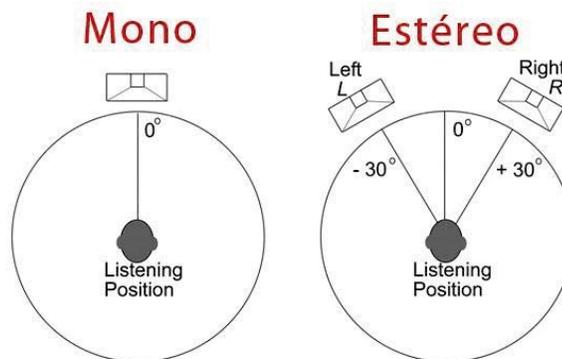


Figura 1-11: Representación básica de sonidos Mono y estéreo [18].

### Multicanal – *Surround*

Otra manera darle espacialidad al sonido, es tener un canal separado para cada dirección deseada. Los sistemas o formatos estéreo multicanal, conocidos también como de sonido envolvente (*surround sound*), tienen muchas variaciones dependiendo del número de canales disponibles a utilizar. Uno de los más conocidos es el sistema 5.1, que se muestra en la figura 1-12.

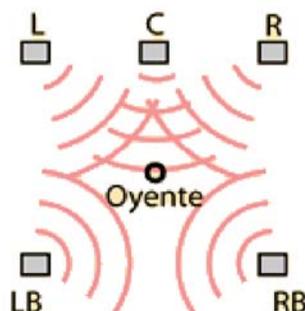


Figura 1-12: Posición óptima de escucha Multicanal [19].

#### 1.4.4 Reconstrucción binaural

Existen modelos de simulación de sonido 3D que no necesariamente requieren múltiples canales para dar una sensación de inmersión, es decir, bastaría con sólo dos canales. La reproducción de señales binaurales a través de los auriculares es una manera que intenta recrear escenas originales para el oyente. El resultado deseado es obtener un sonido espacial tridimensional donde el oyente sea capaz de detectar la fuente de los sonidos en 360°, como si estuviera presente en el ambiente grabado. Para esto, se deben recrear los niveles de sonido en la parte externa de la oreja derecha e izquierda como si el oyente realmente estuviera escuchando una fuente sonora. La grabación se puede realizar mediante la colocación de una cabeza de maniquí en un campo sonoro, pero también puede ser sintetizado en un equipo mediante la síntesis binaural.

## Grabación binaural con maniquí acústico

Las cabezas artificiales, son modelos que intentan reproducir con la mayor aproximación las características auditivas de una cabeza humana real, y para ello cuentan con micrófonos de presión en ambas orejas. Estos maniquís se exponen a una fuente sonora, la cual es grabada por los micrófonos presentes en sus oídos, dando como resultado dos señales correspondientes a las ondas que captura el oído del maniquí. Se obtiene así, una señal binaural que hace creer al oyente que los sonidos vienen de fuentes externas, cuando realmente son señales grabadas en un entorno acústico grabado. En la figura 1-13 se puede apreciar un diagrama esquemático de este procedimiento de grabación.

Cabe mencionar que este experimento no sólo podría hacerse con una grabación y posterior reproducción en los audífonos del oyente. También podría realizarse en tiempo real, esto quiere decir, que podrían captarse los sonidos a través de los micrófonos del maniquí y estos serían transmitidos en tiempo real hacia el oyente a través de los audífonos.

Uno de los principales inconvenientes de usar maniquís acústicos radica en que la función HRTF cambia con cada individuo, de manera que las sensaciones que se pueden percibir con una grabación realizada con un maniquí, pueden ser distintas a la del oyente obteniendo percepciones confusas. Se sugiere que para una representación perfecta de sonido 3D mediante auriculares, se mida cada una de las HRTF de los oyentes. Sin embargo, debido a la imposibilidad práctica de medir HRTF de cada oyente, en la actualidad la tendencia se sitúan en la línea de disponer de un conjunto de HRTF relativamente amplio y aplicar el más adecuado al oyente que va a utilizar el sistema, en función de algún aspecto físico sencillo de su anatomía, como puede ser la distancia interaural, el tamaño de la cabeza, el tamaño de la oreja, etc.

Se puede decir entonces, que la señal binaural es aquella que ya ha sido transformada y por tanto incluye alguna HRTF.

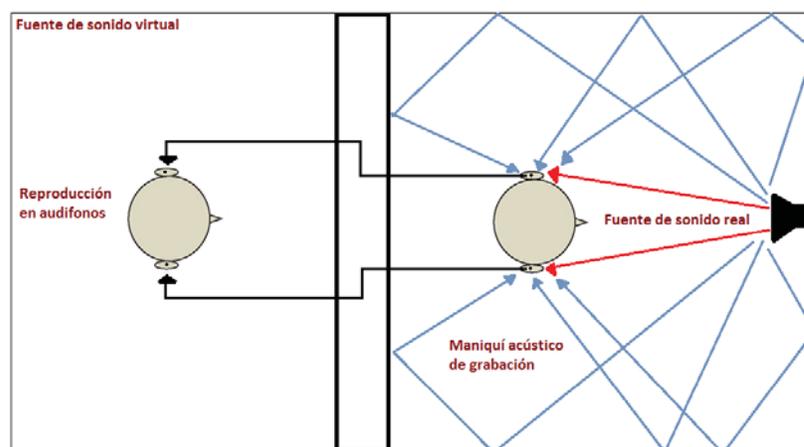


Figura 1-13: Grabación binaural con un maniquí *dummy head*.

## Síntesis binaural a través de convolución

A pesar de la complejidad del estudio de las HRTF, estas ya han sido utilizadas para sintetizar sonidos a través de audífonos. Así, suponiendo que el sistema auditivo asocia estos valores espectrales a una localización espacial particular, la HRTF se pueden usar para filtrar un sonido monoaural en un sonido binaural que sonará como si se originara el sonido en un lugar determinado. Esto se puede lograr usando un sistema de convolución, función matemática que básicamente, opera la señal entrante con una señal de la HRTF medida en el punto donde se quiere localizar el sonido. En la figura 1-13, se presenta un diagrama de bloques explicativo del proceso de síntesis binaural a través de convolución.

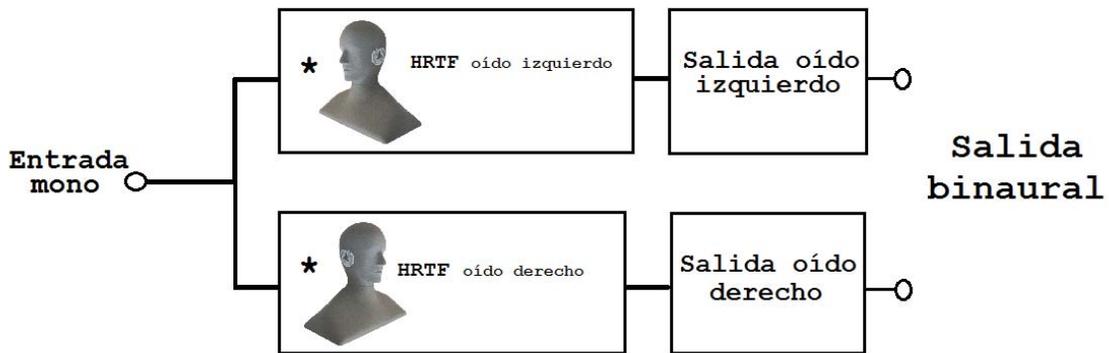


Figura 1-13: Diagrama de bloques de síntesis binaural.

## Reproducción binaural

Una vez que las señales binaurales han sido creadas, se debe encontrar la manera de que sean entregadas de manera independiente y directa a los oídos del oyente, para evitar añadir otra HRTF contraproducente. Esto se hace típicamente a través de audífonos, sin embargo, de igual manera se presentan algunos problemas que hay que tener presentes.

Existe un efecto de localización dentro de la cabeza, que hace que las imágenes perceptivas de los sonidos se ubiquen distorsionadas en los ángulos acimutales y de elevación, resultando más cercanas de lo que en la realidad están (Ver figura 1-14). Sumado a esto, el acoplamiento del auricular al pabellón auditivo crea irregularidades frecuenciales, afectando la posición normal de la pinna. Para poder mitigar estos problemas se han diseñado filtros que alisan la respuesta, no obstante, se recomienda contar con la función de transferencia del audífono (HpTF, por sus siglas en inglés), para poder contrarrestar cualquier efecto que pueda modificar la experiencia virtual o la percepción del sonido.

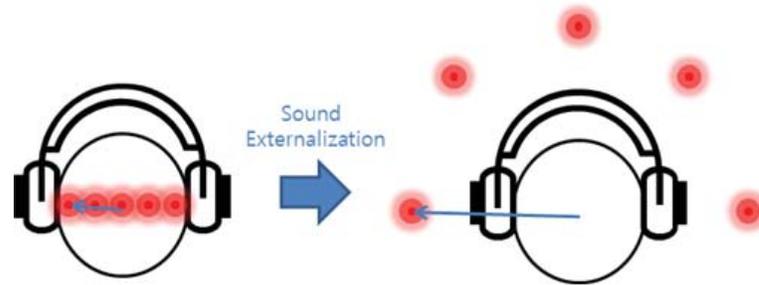


Figura 1-14: Representación gráfica de efecto “in head”. [20]

Otra desventaja de este método es que la escena virtual solo será válida si el oyente mantiene su cabeza fija en una posición conocida. A la hora que se incluye un movimiento de cabeza, las señales binaurales se moverán con ella y dejará de parecer un efecto real. El seguimiento de la cabeza (*Head Tracking*) es la tecnología que permite rastrear la posición de la cabeza del oyente y, dependiendo de una referencia, actuar modificando la síntesis binaural. Estos cambios son medidos por un dispositivo, entregando la orientación y coordenadas específicas del movimiento que realiza la cabeza. Cada vez que éste cambia de orientación, el dispositivo receptor actualiza la posición de la fuente sonora debido a las perspectivas que se obtienen desde distintas posiciones.

## 1.5 Solución propuesta

Si bien, se han realizado variadas investigaciones sobre la simulación de audio 3D, así como también se han desarrollado numerosos experimentos de localización de audio 3D con buenos resultados. En los casos en que se usa una HRTF generalizada, es decir cuando no es la HRTF que representa el canal auditivo del oyente con el que se está ejecutando el experimento, no se logrará tener una sensación completamente real de escucha. Esto se debe a que la estructura externa del oído y las características físicas del cuerpo humano son diferentes para cada persona, lo cual vuelve más difícil la realización de simuladores de audio 3D.

Estas modificaciones que realiza la estructura del oído a la señal acústica, están definidas como funciones de transferencia relacionada con la cabeza (HRTF), las cuales también varían según la posición en el entorno que se encuentre. Por lo tanto resulta complicado construir un sistema estándar de simulación de audio que funcione de la misma manera para todas las personas y replicar el 100% de la información auditiva captada en la realidad.

Estimar completamente la posición de una fuente sonora requiere de varios parámetros como el ITD, el ILD, la respuesta del pabellón auditivo, el movimiento de la cabeza y el movimiento de la fuente, las características de la distancia a la fuente, la reverberación y el eco. Generalmente, se le da una mayor importancia a la ITD y la ILD porque fisiológicamente, a la hora de localizar fuentes, el ser humano es más hábil estimando el acimut, medianamente hábil estimando la elevación y poco hábil determinando la distancia a la fuente.

Hasta ahora, se han realizado algunos sistemas de reproducción de sonido que simulan un espacio tridimensional, donde destacan altavoces estereofónicos, reconstrucción binaural,

entre otros. Todos estos cumplen con recrear la sensación de audio 3D, pero con determinadas limitaciones. En efecto, se busca implementar un sistema donde la reproducción de audio combine las modificaciones causadas por las diferencias fisiológicas humanas y las distintas ubicaciones de las fuentes sonoras, y a su vez con la capacidad de captar toda la información de las señales acústicas que llegan a los oídos.

Los altavoces sitúan los lugares con mayor precisión que los auriculares, ya que, efectivamente, los sonidos vienen desde puntos reales definidos en el espacio, en comparación al auricular donde los sonidos son generados a pocos centímetros del pabellón auditivo y de forma virtual. Sin embargo, los auriculares facilitan de gran forma la creación de una imagen sonora posterior y superior, ya que para realizarlo con parlantes se necesitaría un arreglo de parlantes considerable en comparación de los audífonos. Por otro lado, los auriculares han tenido algunos problemas para reproducir las imágenes sonoras frontales utilizando HRTF no individualizadas, ya que las imágenes sonoras tienden a ser localizadas muy cerca de la cabeza (*Efecto In Head*) [20]. Los auriculares son muy eficaces cuando se utilizan con una HRTF individualizada, donde la respuesta a localización es muy fiel en comparación a las condiciones en un campo sonoro libre [21]. No obstante, también se puede construir un sistema de altavoces utilizando HRTF individualizada.

Se deben hacer pruebas que permitan decidir qué tan importante es usar una HRTF individual o estándar. Además, de estudiar cómo implementar un seguidor del movimiento de la cabeza (*Head tracking*), con la finalidad de implementar un sensor de movimiento. Estas herramientas permitirán detectar el movimiento de la cabeza y poder asociarlo a una HRTF correspondiente, los cuales se convolucionaran en tiempo real, tratando de dar la percepción tridimensional adecuada a la posición y orientación de la cabeza.

Este trabajo en particular, se centrará en lo que será el procesamiento de las mediciones de HRTF para poder implementar una simulación de audio 3D con convolución en tiempo real y con seguimiento del movimiento de la cabeza (*Head tracking*). La finalidad de esto es poder sentar bases para la realización de un experimento de localización con el que se puedan comparar y evaluar los diferentes desempeños de las HRTF genéricas e individualizadas, así como la comparación con sistemas estereofónicos multicanales.

Considerando todo este contexto, es que se plantea el siguiente objetivo general para este trabajo:

## 1.6 Objetivo general

- Realizar la preparación de *setup* necesario para realizar experimentos psicoacústicos de localización de audio 3D, con convolución de HRTF(s) en tiempo real, utilizando sensores de movimiento.

Para cumplir con este objetivo general, se plantean los siguientes objetivos específicos:

## 1.7 Objetivo específicos

- Comprender las técnicas humanas de localización de sonido.
- Investigar sobre técnicas de reproducción de sonido 3D.
- Entender el uso de las funciones de transferencia asociadas a la cabeza (HRTF).
- Estudiar sobre formatos de almacenamiento de audio direccional.
- Investigar sobre softwares de auralización.
- Realizar experimentos de simulación de audio 3D.
- Realizar comparaciones de mediciones propias con bases de datos externas.
- Obtener mediciones de HRTF propias (LAV-EIE, PUCV).
- Hacer pruebas simuladas que permitan evaluar el desempeño de las HRTF.
- Realizar comparaciones de las HRTF obtenidas con otras bases de datos.
- Realizar montaje de sala de pruebas.

## 2 Metodología

En este apartado se explicará sobre algunas herramientas que serán vitales para plantear una simulación 3D, así como algunos formatos de almacenamiento de audio direccional que permita almacenar las mediciones de HRTF realizadas.

### 2.1 Virtual Acoustic

Virtual Acoustics (VA) es un *software* creado en el Instituto de Acústica Técnica (ITA) en la RWTH Aachen University (Alemania), como aplicación de auralización. Permite la auralización en tiempo real, proporcionando módulos e interfaces para experimentos y demostraciones, es de código abierto y totalmente controlable. Gracias a varias contribuciones a lo largo del tiempo en el ámbito de tesis de licenciatura, maestría y doctorado, VA se ha reinventado y mejorado para cumplir con los estándares de vanguardia en la reproducción de audio y reproducción de audio espacial [22].

VA crea un sonido audible desde una situación puramente virtual. Para hacerlo, puede utilizar los datos de entrada digitales pregrabados, medidos, modelados o simulados. Sin embargo, VA crea mundos auditivos dinámicos que pueden ser tratados interactivamente, ya que explica las modificaciones de la situación virtual. En el caso más simple, esto significa que las fuentes de sonido y los oyentes se pueden mover (según las limitaciones de diseño), modificando, en consecuencia, el sonido. Este enfoque de auralización en tiempo real sólo puede lograrse si ciertas partes del procesamiento de audio se actualizan de forma continua y rápida. El resultado es una transmisión de audio que representa la situación, pero de forma virtual.

En VA se encuentran tres entidades virtuales diferentes que representan objetos de sonido. VA utiliza el término receptor de sonido en lugar de oyente. La razón es que los oyentes reducirían la entidad receptora a las criaturas vivas, mientras que en VA esos oyentes también pueden ser, por ejemplo, micrófonos virtuales.

VA diferencia entre la fuente de sonido y la señal de sonido. Una fuente de sonido es una entidad acústica que emite sonido, por otro lado, una señal de sonido que representa la información acústica emitida. Por lo tanto, una fuente de sonido siempre está conectada con una fuente de señal. Por ejemplo, un piano es una fuente de sonido, la música que se reproduce cuando se utilizan sus teclas se denomina fuente de señal.

### 2.1.1 Directividades (HRTF y HRIR)

El patrón de radiación de sonido de una fuente de sonido generalmente se describe mediante una función direccional, que depende de la longitud o frecuencia de la onda. Esta función generalmente se llama directividad. Éstas varían según la fuente de señal y se obtienen desde datos simulados, medidos en direcciones muestreadas (en una rejilla esférica regular o irregular). La directividad de la fuente de sonido describirá la forma en que se propagan las ondas de sonido en el espacio. Por ejemplo, la directividad de una trompeta no es la misma que la de un discurso humano. A los receptores de sonido también se les pueden asignar directividades, las que están determinadas por las HRTF.

### 2.1.2 Representación de audio en VA

La versión actual de VA (v2018b) proporciona módulos de representación, en la Tabla 2-1 se presentan algunos módulos de renderizado. En VA, se puede tener tantos módulos de representación como se necesite, incluidas varias instancias de la misma clase. Esto tiene sentido, por ejemplo, si se desean usar diferentes configuraciones y evaluar el resultado al cambiar entre representaciones en una fracción de segundo. Los módulos de renderización están conectados a módulos de reproducción, además un renderizador también puede alimentar múltiples reproducciones. Sin embargo, hay límites en el número de instancias debido a la potencia computacional disponible. En primera instancia y para partir de una representación lo más simple posible, se ocupara el módulo de renderizado “BinauralFreeField”.

Tabla 2-1 Clases de módulo de renderizado de audio actualmente disponibles en VACore [22].

Nombre de la clase	Flujo de salida	Descripción
BinauralFreeField	Binaural de 2 canales	Representación de campo libre binaural que omite cualquier geometría actual. Utiliza filtrado de FIR para HRTF / HRIR, líneas de retardo de variables y bancos de filtros para directividades por par fuente-receptor.
BinauralArtificialReverb	Binaural de 2 canales	Mezcla la reverberación en el lado del receptor utilizando el tiempo de reverberación, el volumen de la sala y el área de la superficie con un enfoque binaural; aplica el efecto usando el filtrado FIR.
BinauralRoomAcoustics	Binaural de 2 canales	Utiliza un programador de simulación de fondo para respuestas de impulso de sala binaural y aplica el efecto por convolución eficiente de filtros de FIR largos por par receptor-fuente.

### 2.1.3 Modo de auralización

VA proporciona un conjunto de fenómenos que se pueden alternar, llamados modos de auralización. Los modos de auralización se pueden controlar de forma global, para cada fuente de sonido y receptor de sonido individualmente. Si un renderizador respectivo considera el modo de auralización dado, el procesamiento correspondiente se habilitará o deshabilitará en función de la combinación lógica de los modos de auralización (sólo si los modos de auralización de fuente, receptor y ajustes globales son positivos, el fenómeno se hará audible).

La mayoría de los modos de auralización sólo son efectivos para ciertos módulos de renderizado y no tienen sentido para otros. Por ejemplo, un renderizador de campo libre (binaural free field) solo expondrá el sonido directo, la directividad de origen y los cambios de efecto Doppler. Todos los otros fenómenos son descartados [22].

### 2.1.4 Reproducción de audio en VA

La versión actual de VA proporciona distintos módulos de reproducción, como se muestra en la Tabla 2-2. En VA, se puede instanciar tantos módulos de reproducción como se necesite, incluidas varias instancias de la misma clase. Esto tiene sentido, por ejemplo, si se desean usar diferentes configuraciones y evaluar el resultado, cambiando entre reproducciones en una fracción de segundo. Un módulo de reproducción puede reenviar la transmisión de audio final a cualquier número de salidas, si los canales físicos coinciden (por ejemplo, cuatro pares de auriculares adicionales). Sin embargo, hay límites en el número de instancias debido a la potencia computacional disponible. Para los módulos de reproducción, se debe estudiar la posibilidad de ocupar la representación “*headphones*”, donde se acopla una función de transferencia asociada al audífono, para eliminar el efecto de los audífonos en la percepción del sonido. De esto último se entrará en detalles más adelante.

Tabla 2-2 Clases de módulo de reproducción de audio actualmente disponibles en VACore [22].

Nombre de la clase	Flujo de entrada	Flujo de salida	Descripción
Talkthrough	Secuencia basada en canales	Variable	Reenvía la transmisión entrante directamente al <i>hardware</i> de audio. Se usa mucho para la reproducción de audífonos llanos y para las representaciones basadas en canales para las configuraciones de los altavoces.
Headphones	Cualquiera de dos canales	Dos canales ecualizados	Reenvía la secuencia entrante después de aplicar la desconvolución FIR, para la ecualización de auriculares si HpTF está disponible.
BinauralMixdown	Cualquier canal basado en binaural de dos canales	Dos canales binaurales	Utiliza la tecnología binaural dinámica con filtrado FIR para simular la reproducción de sonido basada en canales desde una configuración de altavoz virtual.

Los módulos se pueden habilitar (Enable=True) o deshabilitar (Enable=false) directamente desde un archivo de configuración ubicado dentro de la carpeta de VA (Figura 2-1).

```

80  [Debug]
84  [Audio_driver]
114 [Calibration]
133 [HomogeneousMedium]
142 [Renderer:MyBinauralFreeField]
159 [Renderer:MyAmbientMixer]
166 [Renderer:MyBinauralArtificialReverb]
181 [Renderer:MyBinauralRoomAcoustics]
199 [Renderer:MyMonauralFreeField]
212 [Renderer:MyHearingAidRenderer]
228 [Renderer:MyGenericRenderer]
238 [Renderer:MyAirTrafficNoiseRenderer]
244 [Renderer:MyVBAPFreefield]
251 [Renderer:MyDummyRenderer]
273 [Reproduction:MyTalkthroughHeadphones]
285 [Reproduction:MySubwooferMixer]
286   Class = LowFrequencyMixer
287   Enabled = false
288   Name = Generic low frequency (subwoofer) loudspeaker mixer
289   Outputs = Cave_SW
290   MixingChannels = ALL
291
292 [Reproduction:MyHD600]
303 [Reproduction:MyNCTC]
318 [Reproduction:MyAmbisonics]
326 [Reproduction:MyMixdownHeadphones]
335 [Reproduction:HearingAidTalkthrough]
342 [Setup]
349 [Output:MyDesktopHP]
353 [Output:MyDesktopLS]
360 [Input:MyDesktopMike]

```

Figura 2-1: Archivo de configuración de VA "VACore.ini".

## 2.2 Formatos.

Las HRTF han sido medidas por varios laboratorios, típicamente se almacenan en el formato de archivo nativo o preferido de cada laboratorio. Si bien los diferentes formatos pueden ser una ventaja para cada investigador, un intercambio de tales datos es difícil debido a la incompatibilidad entre formatos [23]. Se han propuestos algunos formatos, que han sido estandarizados, para almacenar HRTF con un enfoque en la intercambiabilidad y extensibilidad.

### 2.2.1 OpenDaff (daff)

OpenDAFF (Open Directional Audio File Format) es un paquete de software libre y de código abierto para contenido de audio direccional como directividades de fuentes de sonido (altavoces, instrumentos musicales) y receptores de sonido (micrófonos, funciones de transferencia relacionadas con la cabeza, HRIR / HRTF). La idea fundamental es almacenar toda la información direccional, básicamente todos los registros son en una grilla esférica, dentro de un solo archivo “.daff” [24].

Para cumplir con una amplia gama de aplicaciones, OpenDAFF admite varios tipos de contenido:

1. Respuestas de impulso (IR).
2. Espectros discretos de Fourier (funciones de transferencia).
3. Espectros de magnitud / fase (frecuencias discretas).

Los tipos 1-2 se relacionan con señales/sistemas y se definen a través de un soporte igualmente espaciado (tiempo o frecuencia). La categoría 3 se refiere a medidas relacionadas con la energía, que se definen para ciertas bandas de frecuencia. Aparte de esto, los datos se pueden definir para un número arbitrario de canales individuales, por ejemplo, en una HRTF hay canales para cada registro, uno para el oído izquierdo y otro para el derecho.

Este formato se inventó inicialmente para proporcionar un acceso extremadamente ligero al contenido. Debido a esto, el formato de archivo, actualmente, sólo admite conjuntos de datos conectados a una grilla esférica regular con posiciones equiangulares en una esfera unitaria (como el sistema de coordenadas globales con latitudes, longitudes, polo norte y polo sur, también denominado cuadrícula de Gauss) (Ver figura 2-2). Esta decisión de diseño permite la búsqueda de un dato vecino más cercano en una esfera, usando solo operaciones de redondeo que permite el uso de la API OpenDAFF, por ejemplo, en la auralización en tiempo real para aplicaciones de realidad virtual, donde la tecnología binaural juega un papel clave [24].

Para resolver las ambigüedades con orientaciones de objetos vs puntos de datos, OpenDAFF presenta dos vistas separadas que se muestran en la figura 2-2. Coordenadas esféricas de datos (DSC) y coordenadas esféricas de objeto (OSC). Para escribir los datos se usa DSC, son el marco de referencia para las direcciones de los registros y no tienen relación con el objeto en sí. Nombramos los ángulos  $\alpha$  (alfa, A) y  $\beta$  (beta, B).  $A_0^\circ$  es el primer meridiano (rojo) y  $B_0^\circ$  marca el polo sur (verde).

Las indicaciones de OSC se refieren al objeto y se usan para leer el archivo daff. Se llaman acimut ( $\phi$ ,  $j$ ,  $P$ ) de  $-180^\circ$  a  $180^\circ$  y elevación ( $\theta$ ,  $J$ ,  $T$ ) de  $-90^\circ$  a  $90^\circ$ . Se definen esos puntos ( $P0^\circ$ ,  $T0^\circ$ ) al frente, ( $P90^\circ$ ,  $T0^\circ$ ) a la izquierda y  $T90^\circ$  apunta hacia arriba. Ambas vistas están mapeadas entre sí mediante rotaciones tridimensionales (convención de inclinación y balanceo de cabeceo) dentro de las coordenadas esféricas. Esta orientación tiene que ser determinada solo una vez y está escrito en el archivo DAFF cuando es creado. Los usuarios preferirán el OSC, porque inequívocamente define las direcciones con respecto al objeto [25].

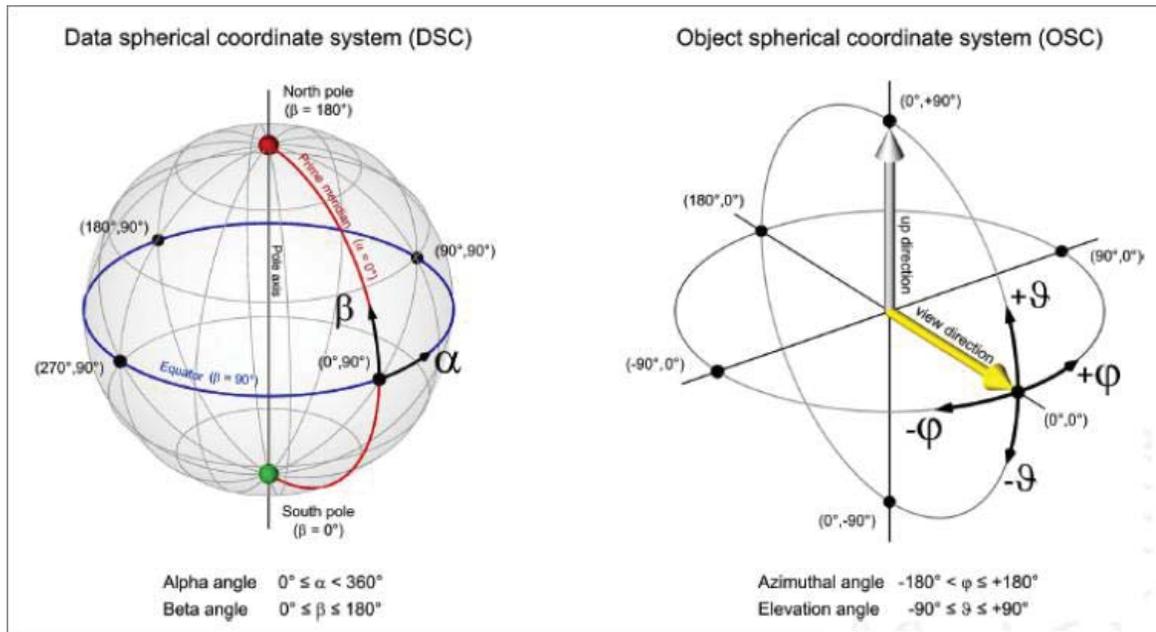


Figura 2-2: Coordenadas esféricas de datos (DSC) y coordenadas esféricas de objeto (OSC) [25].

OpenDaff se complementa con herramientas útiles, como un visor gráfico para DAFF datos de audio direccionales (DAFFViewer, figura 2-3). Éste permitirá explorar los archivos de mediciones de HRTF que se hayan realizado o algunas mediciones pertenecientes a otros laboratorios.

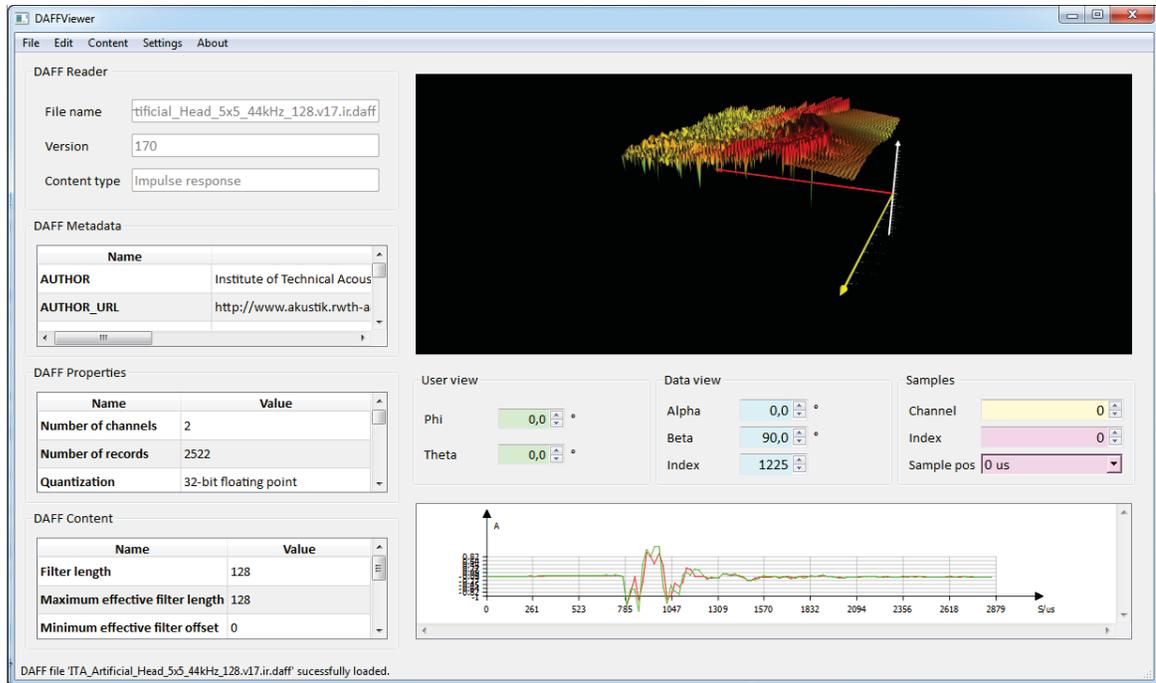


Figura 2-3: DAFViewer.

### 2.2.2 SOFA

Es un formato de archivos para almacenar o representar datos acústicos orientados espacialmente, permitiendo almacenar no sólo HRTF o IR (Impulse Response), sino también datos más complejos, como por ejemplo, respuestas direccionales de impulso de sala (DRIR), medidas con un conjunto de micrófonos multicanal, excitados por un arreglo de altavoces [26].

SOFA nace como propuesta de formato para almacenar datos acústicos con un enfoque en la intercambiabilidad y extensibilidad. Este formato ha sido estandarizado por la Audio Engineering Society (AES) como AES69-2015 [27].

SOFA tiene como objetivo cumplir los siguientes requisitos:

- Descripción de una configuración de medición con geometría arbitraria, es decir, no limitada a casos especiales como una cuadrícula regular, o una distancia constante.
- Datos de auto descripción con una definición coherente, es decir, toda la información requerida sobre la configuración de medición debe proporcionarse como metadatos en el archivo.
- Flexibilidad para describir datos de múltiples condiciones (oyentes, distancias, etc.) en un solo archivo.
- Soporte parcial de archivos y redes.
- Disponible como archivo binario con compresión de datos para un almacenamiento y transferencia eficientes.
- Convenciones de descripción predefinidas para las configuraciones de medición más comunes.

En pocas palabras, la configuración de medición se describe por varios objetos y sus relaciones. Se considera una medición como una observación discreta muestreada, realizada en un momento específico y bajo una condición específica. Cada medición consta de datos (por ejemplo, una respuesta de impulso, IR) y se describe por sus dimensiones y metadatos correspondientes. Las convenciones de SOFA dan una descripción consistente de las configuraciones de medición, es decir, dejan explícitas la configuración de medición que se utilizó [26].

### 2.2.3 Especificaciones generales y convenciones

Para cumplir con los diferentes requisitos provenientes de diferentes campos de aplicación, se proponen convenciones SOFA, es decir, definiciones de datos y metadatos consistentemente describiendo configuraciones de medición de HRTF.

#### Objetos

El oyente es el objeto que incorpora todos los receptores. Para HRTF, un oyente puede ser, por ejemplo, un micrófono en el canal auditivo de un maniquí. Un oyente representa la estructura del micrófono como una esfera. Incorporar a los receptores en el oyente como un único objeto es importante, porque en las mediciones, generalmente la orientación y/o la posición del oyente varían sin cambios sustanciales en la relación cabeza-micrófono. Por ejemplo, en mediciones hechas para múltiples posiciones en una habitación, la posición de la cabeza varía y la relación entre la cabeza y los micrófonos no cambia. Por lo tanto, es fundamental tener en cuenta que solo se considera un oyente.

Emisor es cualquier excitación acústica utilizada para la medición. Las cantidades de emisores no están limitadas en SOFA. La contribución del emisor particular se describe por los metadatos, posteriormente se detallara más sobre los metadatos. La fuente es el objeto que incorpora todos los emisores. En SOFA, una fuente puede ser un altavoz controlado, una matriz de altavoces (con los hablantes particulares como emisores), un coro (con el humano en particular como emisor), etc. Se debe tener en cuenta que sólo una fuente se considera, pero la fuente puede incorporar un número ilimitado de emisores.

La habitación es el volumen que encierra la configuración de la medición. En el caso de una medición de campo libre, la habitación no es considerada. La descripción de habitación opcional para mediciones realizadas en espacios reverberantes, con una descripción directa de una simple caja de zapatos (*shoebox*), o con un enlace a un archivo de intercambio de activos digitales para una descripción más compleja [28].

#### Relación entre los objetos

Se usarán dos sistemas de coordenadas, el sistema de coordenadas global y el sistema de coordenadas local, ambos en coordenadas cartesianas o esféricas.

El sistema de coordenadas global definirá la posición de los objetos fuente y oyente dentro de la habitación, los que compartirán el mismo sistema coordinado. El origen del sistema de coordenadas global se puede elegir arbitrariamente.

Los sistemas de coordenadas locales definirán las posiciones de los emisores dentro del objeto fuente, mientras que las posiciones de receptores lo harán dentro del objeto oyente. Todos los emisores compartirán la misma fuente relacionada local sistema coordinado; todos los receptores compartirán el mismo sistema de coordenadas local relacionado con el oyente [28]. La relación de los sistemas de coordenadas locales y globales se muestra en la figura 2-4.

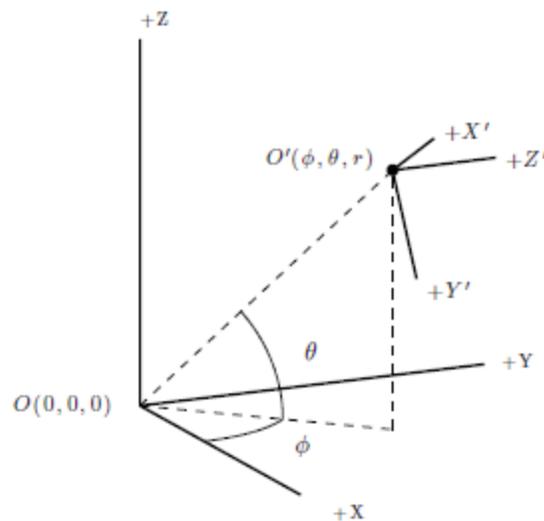


Figura 2-4: Sistemas de coordenadas local y global [27].

El origen del sistema de coordenadas local está en la posición  $(\phi, \theta, r)$ , relativa al sistema de coordenadas global, que en este ejemplo está en coordenadas esféricas. La orientación del sistema de coordenadas local está dado por los vectores de “vista” y “arriba”, que definen la dirección del eje X 'positivo y eje Z', respectivamente [27].

#### **Convención de medición de HRTF en campo libre (SimpleFreeFieldHRTF)**

Las mediciones se hacen a distancia constante en el campo libre, con una sola excitación fuente suponiendo un altavoz omnidireccional. Se consideran oyentes humanos y, por lo tanto, la cantidad de receptores son dos, correspondiente a los dos canales auditivos. Los ángulos de acimut y elevación son variados y la inclinación de la cabeza no se considera durante la medición. Los HRTF medidos se representan como filtros, con un solo conjunto de HRTF de un oyente por archivo.

La convención SOFA SimpleFreeFieldHRTF define la configuración mostrada en la Figura 2-5, sobre el planteamiento de una escena de medición acústica. Para mayores detalles, consultar AES69-2015 [26].

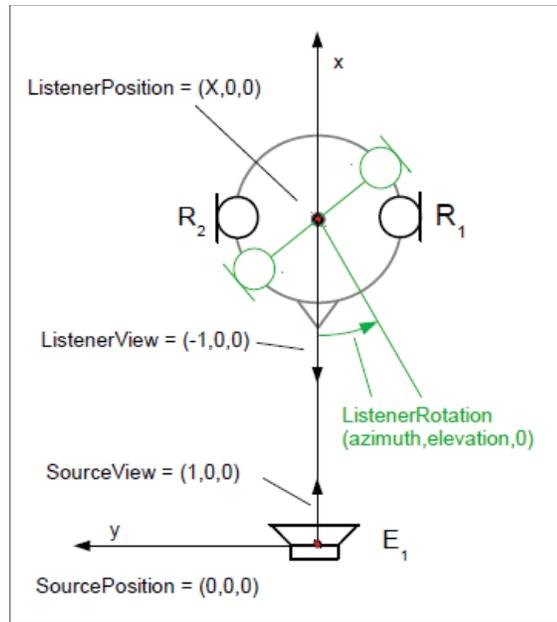


Figura 2-5: Configuración de medición de HRTF de campo libre considerada en la convención SOFA "SimpleFreeFieldHRTF" [28].

## 2.3 ITA Toolbox para Matlab

En este apartado se explicará sobre otra herramienta (Toolbox) para Matlab, que será de ayuda para procesar mediciones y adecuarlas a los formatos necesarios.

ITA Toolbox es una herramienta de Matlab, utilizada para resolver tareas comunes de procesamiento en el campo de la investigación acústica, incluida la importación y exportación de datos, así como diferentes representaciones gráficas de los datos [29].

Los datos de medición o simulación se almacenan comúnmente de forma numérica en formato vectorial o matricial. Asociado con los datos, está la necesidad de almacenar meta información, (frecuencias de muestreo, comentarios, coordenadas, unidades físicas o de dominio). Esta herramienta es capaz de calcular valores y unidades de multiplicación y división de dos variables, así como transformar, por ejemplo, de coordenadas cartesianas a cilíndricas o esféricas (viceversa). Los objetos almacenan los datos junto con la información del dominio, se almacenan en la clase "itaCoordinates" simplemente utilizando: ".cart" para cartesianos o ".sph" para representación esférica.

Los datos de audio son comúnmente mucho más complicados que los valores individuales o las coordenadas. El mismo concepto se utiliza y amplía hacia una clase llamada "itaAudio", la que almacena datos de audio muestreados equidistantemente a partir de mediciones o simulaciones en tiempo o frecuencia. A los datos de dominio de tiempo siempre se puede acceder por ".time" y los datos de frecuencia por ".freq".

## Funcionalidad

La funcionalidad básica se muestra en la Figura 2-5, dividida en diferentes grupos, con el concepto de OOP (programación orientada a objetos). Los círculos restantes se describen en sentido horario: las rutinas de importación y exportación se suman en dos funciones “ita\_read” y “ita\_write” que son capaces de leer y escribir formatos de datos comunes, tales como, *wav*, *daff*, *sofa*, *mat*, entre otros.

La rutina de importación busca automáticamente metadatos e integra su información en los objetos de audio. Una vez que los datos se almacenan en “itaAudio”, varias rutinas de gráficos están disponibles directamente utilizando, por ejemplo, “.plot\_freq”, “.plot\_time” y “.plot\_spectrogram” para gráficos de frecuencia, tiempo y dominio espectral.

La ventaja de usar archivos .m de ITA-Toolbox reside en los metadatos almacenados junto con los datos de audio. No es necesario especificar la frecuencia de muestreo al filtrar los datos, ya que se lee automáticamente desde el objeto en sí. Se puede acceder a todas las funciones, métodos y propiedades de las clases a través de la línea de comando de MATLAB, con acceso directo a los datos en tiempo y dominio de frecuencia siempre posible sin transformada de Fourier explícita [30].



Figura 2-5: Descripción general de la funcionalidad de ITA-Toolbox para MATLAB [30].

## 3 Simulación de audio 3D

Con el objetivo de entender las capacidades de Virtual Acoustic (VA), se fijó la tarea de generar simulaciones virtuales con diferentes características espaciales, de las cuales se entrará en detalle más adelante.

Como se estudió anteriormente, Virtual Acoustic sólo trabaja con el formato “.*daff*”, por lo tanto, para poder hacer una experimentación y comparación práctica con el formato “SOFA” se utilizaron otros medios proporcionados por Matlab para la auralización y simulación 3D.

Para realizar las simulaciones se utilizaron bases de datos de HRTF medidas por otros laboratorios, con la finalidad de sentar bases para lo que será la experimentación con HRTF individualizadas, medidas por el Laboratorio de acústica y vibraciones de la escuela de ingeniería eléctrica de la Pontificia Universidad Católica de Valparaíso.

Para poder cargar y revisar datos como respuestas impulso y respuestas en frecuencia, dentro de un archivo “DAFF” y/o “SOFA”, revisar apéndice (A1).

### 3.1 Virtual Acoustic – OpenDAFF

En el siguiente apartado, se revisará con detalle los pasos y requerimientos mínimos para realizar una correcta simulación de audio 3D, en Virtual Acoustic. Estas estarán divididas por etapas y por tipo de experimento. Los tipos de experimento que se revisarán serán:

- ✓ Simulación de audio 3D de una fuente de sonido con posición fija en la escena.
- ✓ Simulación de audio 3D de una fuente de sonido con posición en movimiento alrededor del oyente.
- ✓ Simulación de audio 3D de dos fuentes de sonido con posición fija en la escena.

Para poder complementar esta explicación, se adjuntan en el apéndice los códigos de MATLAB correspondientes a cada experimento, además de un pequeño diagrama resumen que describe el proceso a la hora de plantear una simulación acústica en VA (Figura 3-1).

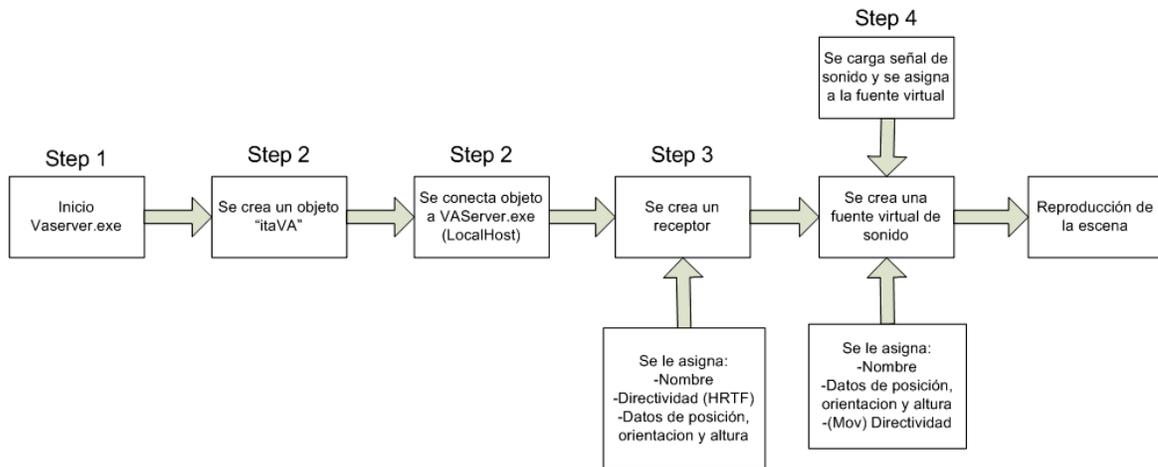


Figura 3-1: Diagrama de proceso para plantear una simulación acústica en Virtual Acoustic.

### Primera etapa: inicio del servidor Virtual Acoustic

Se establece una serie de comandos que dejan la apertura de “VaServer.exe” de forma automática, es decir, una vez que se da inicio a la simulación vía Matlab, automáticamente se abrirá “VAServer” y empezará a cargar los directorios importantes para el programa, los que se muestran dentro de la ventana de comandos (Ver figura 3-2). También se mostrarán algunos datos de la configuración de la simulación, tales como: frecuencia de muestreo, canales de salidas, tipo de renderizador y tipo de reproductor ocupado. Una vez cargados todos los directorios importantes, se dará inicio a la simulación mostrando una serie de comandos para poder manejar datos como ganancia de salida, entre otras cosas.

```

Administrator: C:\Windows\system32\cmd.exe
'Portaudio' driver
[ VAInfo ] Core | Streaming at 44.1 kHz on no inputs and 2 o
utputs with a buffer size of 256 samples.
[ VAInfo ] Core | Started one rendering module
[ VAInfo ] Core | + MyBinauralFreeField (BinauralFree
Field) feeding one reproduction module.
[ VAInfo ] Core | Started one reproduction module
[ VAInfo ] Core | + MyTalkthroughHeadphones (Talkthro
ugh) feeding one output group.
[ VAServer ] Core version: UACore v2018.a (release)
[ VAServer ] Successfully started and listening on 0.0.0:12340
[ VAServer ] Controls:
[ VAServer ] m Toggle output muting
[ VAServer ] + Increase output gain by 3dB
[ VAServer ] 0 Output gain 0dB
[ VAServer ] - Decrease output gain by 3dB
[ VAServer ] r Reset core
[ VAServer ] l Circulate log levels
[ VAServer ] c List connected clients
[ VAServer ] s List search paths
[ VAServer ] q Quit
[ VistaIOHandleBasedIOMultiplexer] Got COMMAND 0
[ VAInfo ] Core | Resetting core
  
```

Figura 3-2: VAServer.exe.

### Segunda etapa: Objeto almacenador de la escena

Se crea un objeto interfaz, por ejemplo, “A” de la clase “itaVA”. Este objeto se conecta al servidor de VA, que se ejecuta dentro del mismo equipo (*localhost*). Esta interfaz es la que almacenará toda la información de la simulación a saber, fuentes de sonido, receptores de sonido, la

geometría subyacente y las directivas del par fuente/receptor. Por ende, es posible modificar la escena virtual a través de esta interfaz.

### **Tercera etapa: Receptor de sonido**

Se crea un receptor y se le asignará una HRIR en formato “daff”. Análogamente se debe establecer la posición y orientación del receptor dentro de la escena, así como su altura relativa a la posición del receptor real (posición de pie o sentada). Por lo general, el receptor estará ubicado siempre en el centro de la escena, ocupando la coordenada  $[0 \ h \ 0]$ , donde  $h$  es la altura del oyente. Hasta el momento, se ha ocupado como altura de referencia 1.2 metros, que sería una altura aproximada de un oyente sentado. Este caso es idéntico en todas las simulaciones que se explicarán (Apéndice A.2- Step 3).

### **Cuarta etapa: Fuente sonora**

Virtual Acoustic nos permite simular varias fuentes de sonido a la vez. Éstas pueden ser de posición fija, es decir, que emiten un sonido constante proveniente de la misma posición, o pueden ser fuentes con posición no fija, es decir, en movimiento (sentido circular alrededor del oyente). En los siguientes apartados se describirá cómo establecer estas fuentes.

Se debe dejar en claro que los procesos anteriormente explicados son idénticos para todos los experimentos que se realizaron. A modo de evitar repeticiones, estas etapas sólo se establecieron en el Apéndice A.2 correspondiente al código de simulación de una fuente fija en el espacio. En el caso del código fuente en movimiento (Apéndice A.3), sólo se especifica el código a partir de la creación de una fuente en movimiento.

#### **Fuente sonora con posición fija (A.2 Step 4)**

Se crea una fuente sonora fija, a la cual se le asigna orientación y posición espacial dentro de la escena virtual. Una vez fijada la posición de la fuente, se le debe asignar una señal a la fuente (en formato WAV). Finalmente, deben agregarse los comandos de reproducción y detención pertinentes. Como se dijo anteriormente, el receptor se dejará fijo en el centro (0,0,0) de la simulación, por lo tanto, la posición o referencia de la posición de la fuente relativa al oyente en coordenadas cartesianas son: a la derecha en X positivo e izquierda X negativo, al frente en Z negativo y atrás en Z positivo, sobre en Y positivo y debajo en Y negativo. Así como en las coordenadas esféricas, están definidas por radio, elevación y acimut (Ver figura 3-3).

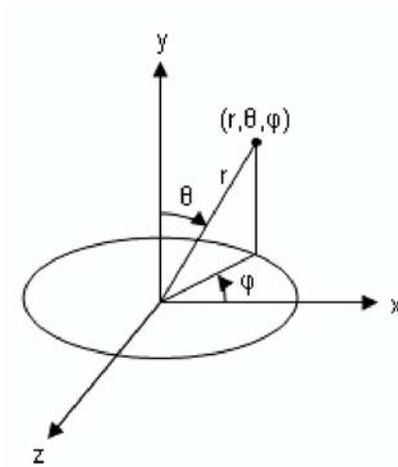


Figura3-3: Eje de referencia para fijar fuente sonora [31].

### Fuente sonora con posición en movimiento (A.3)

Al igual que el caso de la fuente fija, se debe crear una fuente de sonido el cual se acopla a una señal de sonido en formato WAV (línea 13 en A.3). Pero ahora, además, debe asociarse a una directividad de fuente, debido a que debe representar cierta propagación, de acuerdo a la fuente sonora que se desea simular, la cual también se almacena en formato “Daff”.

Para poder fijar la ruta en movimiento de la fuente de sonido (contrario a las manecillas del reloj), se deben fijar los parámetros señalados en la Tabla 3-1.

Tabla 3-1 Parámetros de recorrido de fuente sonora en movimiento en VA.

circleR	Radio de la trayectoria [metros]
nlegs	Número de pasos equidistantes en la trayectoria
Tvel	Tiempo en recorrer los pasos de “nlegs”
phi_start	Punto de inicio en el acimuth [rad]
phi_end	Punto de término en el acimuth [rad]
theta	Ángulo del Zenith [rad]

Al igual que los casos anteriores, se finalizará asignándole una orientación a la fuente y agregando los comandos de reproducción y término de reproducción correspondientes, que para este caso de la fuente en movimiento, va complementado con un “for” que hace que la reproducción del sonido dure hasta que se cumpla el recorrido completo predefinido (Líneas 62 a 68, del A.3).

## Más de una fuente de sonido en la escena

Para trabajar con más de una fuente sonora actuando de forma paralela, se debe agregar una nueva etapa (Etapa 5), que se procederá a describir a continuación.

### Etapa 5: Sincronización de la escena: Dos fuentes sin movimiento (A.3)

Para esta simulación se mantienen los procesos anteriormente explicados. Pero ahora, se debe agregar una etapa de sincronización de las fuentes sonoras, en el cual simplemente se debe posicionar y orientar ambas fuentes dentro de la escena y se deben agregar los comandos de reproducción y detención del sonido asociadas a ambas fuentes.

## 3.2 Grabación de los resultados de la simulación

En Virtual Acoustic, se puede guardar el audio resultante de la simulación. El único requisito es activar la grabación por configuración antes del tiempo de ejecución. Esto se realiza en la zona de configuración del módulo de renderizador y reproducción, es decir, en el archivo de configuración de VA "VACoreOriginal.ini".

Actualmente, se está utilizando un modo de renderizado que es "*BinauralFreeField*" y dos modos de reproducción. Los modos de reproducción que se han ido alternando son "*MyTalkthroughHeadphones*" y "MyHD600". El primero se ocupa para una reproducción convencional con audífonos, en cambio el segundo se ocupa para contemplar la respuesta del audífono durante la simulación, con la finalidad de poder aumentar la sensación de inmersión en la simulación virtual. En otras palabras, se toma en cuenta el efecto del audífono sobre la oreja.

En los casos de "*BinauralFreeField*" y "*MyTalkthroughHeadphones*", se puede activar un modo de grabación (habilitando con un Enable: True/False) especificando tanto el directorio de grabación así como el nombre del archivo en el cual se almacenará la prueba de audio, como se puede observar en la figura 3-4. En la línea 148(Vacore) se encuentra la opción "*RecordOutputEnable*" para habilitar o deshabilitar la grabación de salida, en este caso del *renderizador*. Seguido de esta línea, se encuentra "*RecordOutputFilePath*" donde se debe especificar el directorio de donde se grabará el archivo en formato ".wav", así como el nombre del archivo. En el caso del modo de reproducción, "MyHD600" no posee salida de grabación directa, pero aparte de estos modos particulares de grabación, existe un modo de grabación de la salida global del sistema en la categoría "*Debug*" del archivo "VACoreOriginal.ini". Para casos futuros, se trabajará principalmente con la grabación de la salida global del sistema.

```

142 [Renderer:MyBinauralFreeField]
143 Enabled = true
144 Class = BinauralFreeField
145 #Reproductions = MyHD600
146 Reproductions = MyTalkthroughHeadphones
147 OutputDetectorEnabled = true
148 RecordOutputEnabled = false
149 RecordOutputFilePath = C:\Users\informatica\Documents\Archivos_matlab_tesis\output_save\2807\Renderer_MyBinauralFreeField_prueba_mov.wav
150 HRIRFilterLength = 256
151 MotionModelNumHistoryKeys = 10000
152 MotionModelWindowSize = 0.1
153 MotionModelWindowDelay = 0.1
154 MotionModelLogInputSources = true
155 MotionModelLogEstimatedOutputSources = true
156 MotionModelLogInputListener = true
157 MotionModelLogEstimatedOutputListener = true
158 SwitchingAlgorithm = linear

```

Figura 3-4: VACoreOriginal.ini, configuración externa de VA.

En una simulación con tipo de reproductor “*TalkthroughHeadphones*”, no existen diferencias entre las grabaciones de salida en cualquier punto. Para complementar esta idea, se observa en la figura 3-5 un diagrama donde se especifica con un “REC” en las salidas en que se puede habilitar la grabación. Para los experimentos, cuando se elige el reproductor “*TalkthroughHeadphones*” se puede grabar en cualquier punto, es decir, a la salida del *renderizador*, reproductor o salida global, ya que no existirán diferencias o variaciones, no así al elegir el modo de reproducción “*MyHD600*”, en el cual se puede apreciar el efecto de contemplar la modificación de la función de transferencia, debido a la presión que ejerce el audífono sobre la pinna. Por lo tanto, existirán diferencias al comparar la grabación de salida global con la grabación de salida del Reproductor “*MyHD600*”. En la figura 3-5, se resumen las alternativas de grabación en VA.

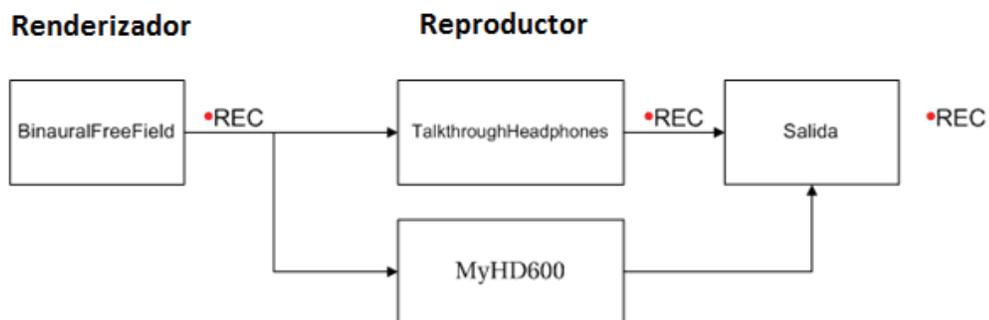


Figura 3-5: Diagrama de posibilidades de grabación de la simulación en VA.

### 3.3 Análisis de resultados de la simulación

Se procederá a analizar resultados obtenidos de algunos experimentos realizados. Se analizarán HRTF medidas para el *Dummy Head* del laboratorio de Acústica y Vibraciones, así como resultados de mediciones obtenidos por otros laboratorios. Se explicará cómo trabajar con los archivos que almacenan los datos de las HRTF (DAFF y SOFA) a través de Matlab.

#### Fuente sonora con posición fija

Para poder visualizar y analizar de mejor forma, se procedió a hacer simulaciones de fuentes virtuales utilizando un impulso sonoro de corta duración, en ubicaciones representativas a la hora de localizar un sonido, como lo son la zona frontal y en las zonas laterales a los oídos. En

otros términos, se realizó una medición de la respuesta impulsiva (IR), para zonas que son determinantes o de fácil interpretación, a la hora de analizar su IR.

Primero, se generó una fuente virtual al lado izquierdo del oyente, tal como se representa en la figura 3-6. El resultado de este experimento simulado se presenta en la figura 3-7, en el que se puede observar cómo la excitación de mayor amplitud correspondiente al canal derecho (rojo), demostrando un efecto acorde a lo esperado, mientras que en la respuesta en frecuencia (Figura 3-8) el efecto sombra se va haciendo cada vez más notorio a medida que aumenta la frecuencia.



Figura: 3-6: Esquema explicativo del experimento simulado. Fuente sonora en la zona lateral.

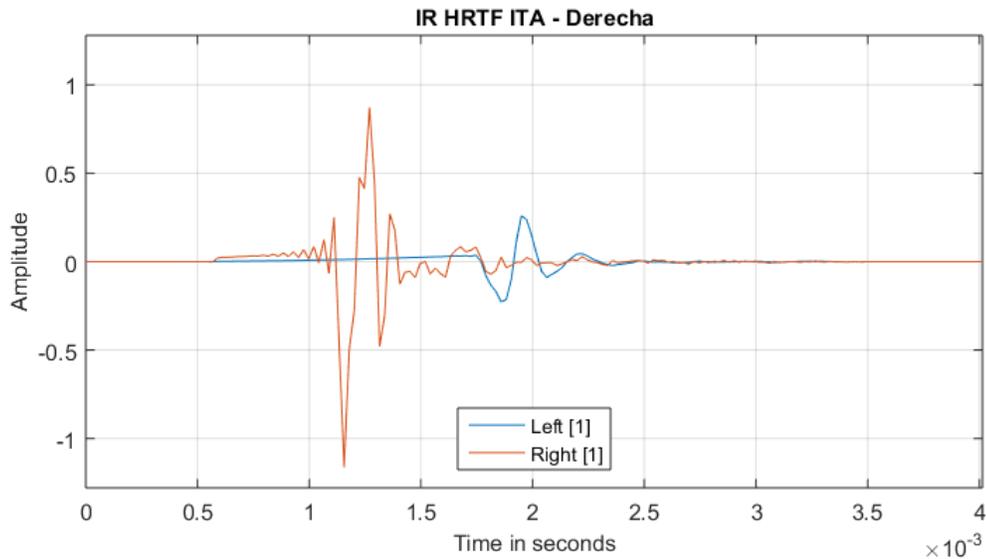


Figura 3-7: Respuesta impulso de HRTF ITA para una fuente sonora ubicada a la izquierda del oyente.

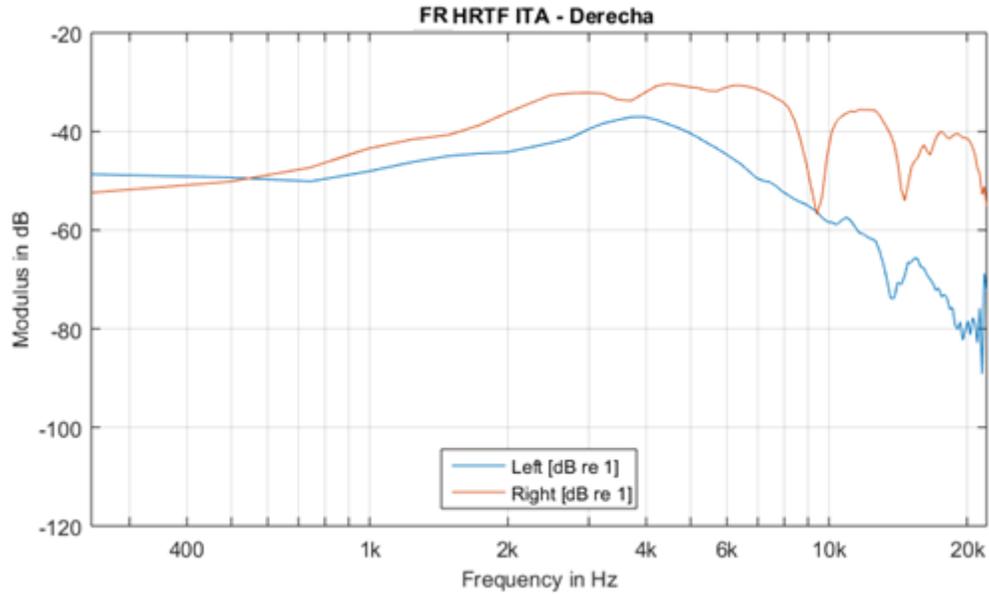


Figura 3-8: Respuesta en frecuencia de HRTF ITA para una fuente sonora ubicada a la izquierda del oyente.

En el caso de una fuente sonora en la zona frontal al oyente, se puede apreciar en la figura 3-10 como ambos pulsos sonoros llegan con amplitud similar y al mismo instante de tiempo para ambos canales. De forma similar, se puede apreciar una respuesta en frecuencia (figura 3-11) prácticamente similar en nivel y forma, a excepción de pequeñas diferencias en altas frecuencias.

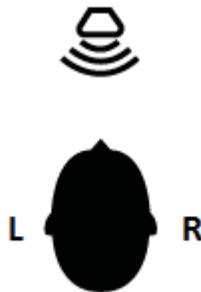


Figura: 3-9: Esquema explicativo del experimento simulado. Fuente sonora en la zona frontal.

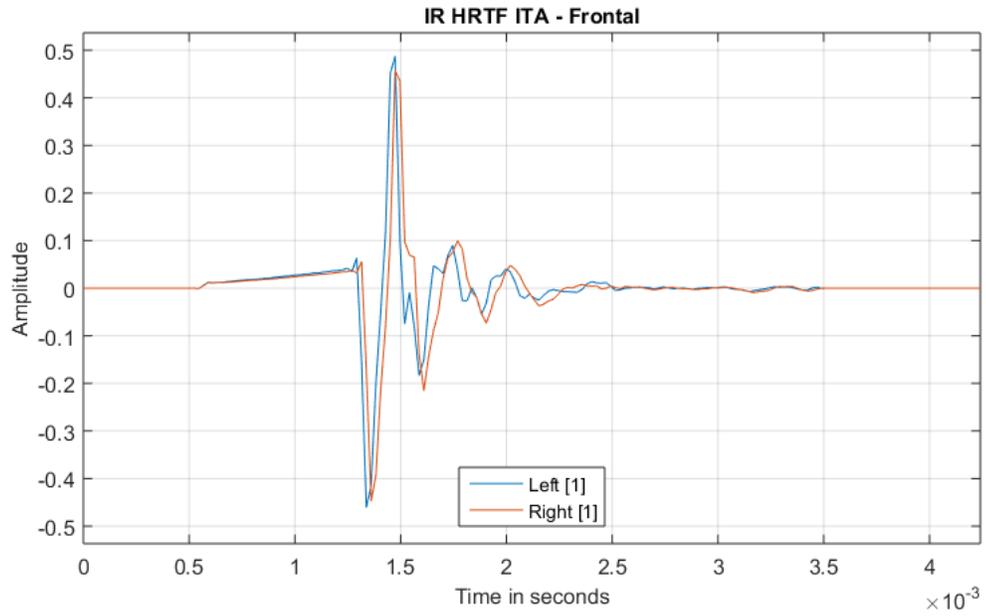


Figura 3-10: Respuesta impulso de HRTF ITA para una fuente sonora ubicada a la frente del oyente.

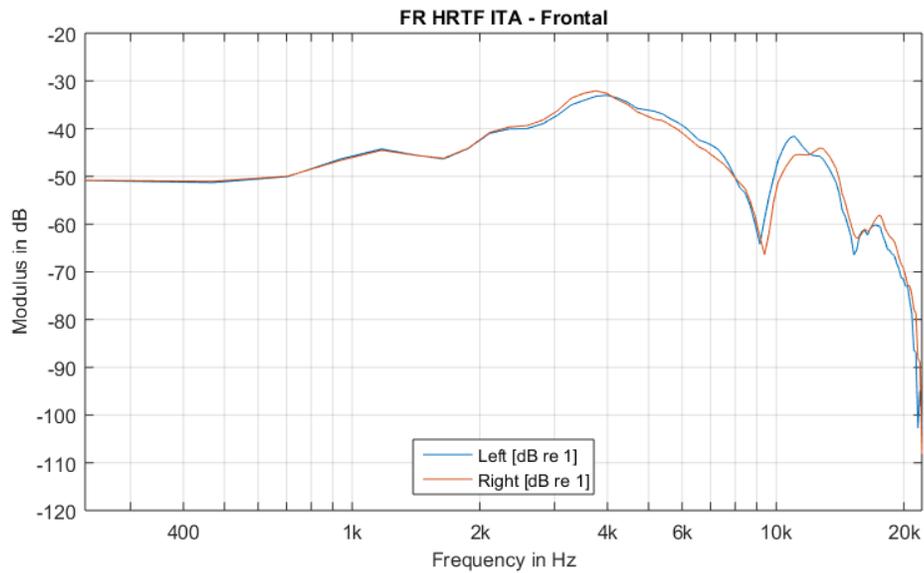


Figura 3-11: Respuesta en frecuencia de HRTF ITA para una fuente sonora ubicada al frente del oyente.

### Fuente sonora con posición en movimiento

Para el caso de la fuente sonora en movimiento, se estableció un recorrido desde el lado derecho del receptor hacia el lado izquierdo. Como el sentido de movimiento es anti horario, la fuente pasará por delante del receptor. Se graba el resultado de la respuesta a un impulso, dando como resultado la figura 3-12, donde se puede apreciar cómo llega el pulso primero al canal derecho (azul) y éste va disminuyendo en amplitud, así como análogamente va aumentando la magnitud del canal izquierdo a medida que éste avanza en esa dirección. Análogo a esto, se adjunta en la figura 3-13 un zoom de la respuesta impulso del primer, cuarto y último impulso, es decir los pulsos del segundo 1, 5 y 10, donde se puede apreciar un efecto similar al explicado con una fuente fija.

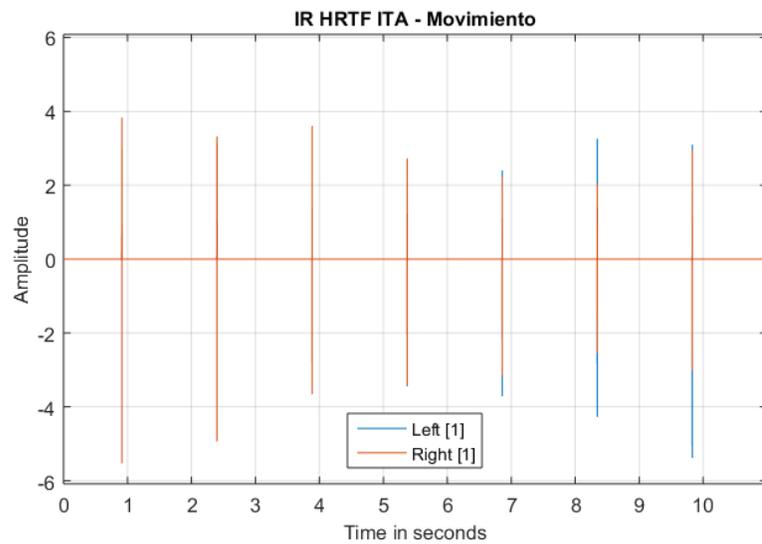


Figura 3-12: Respuesta de HRTF ITA a un impulso en movimiento.

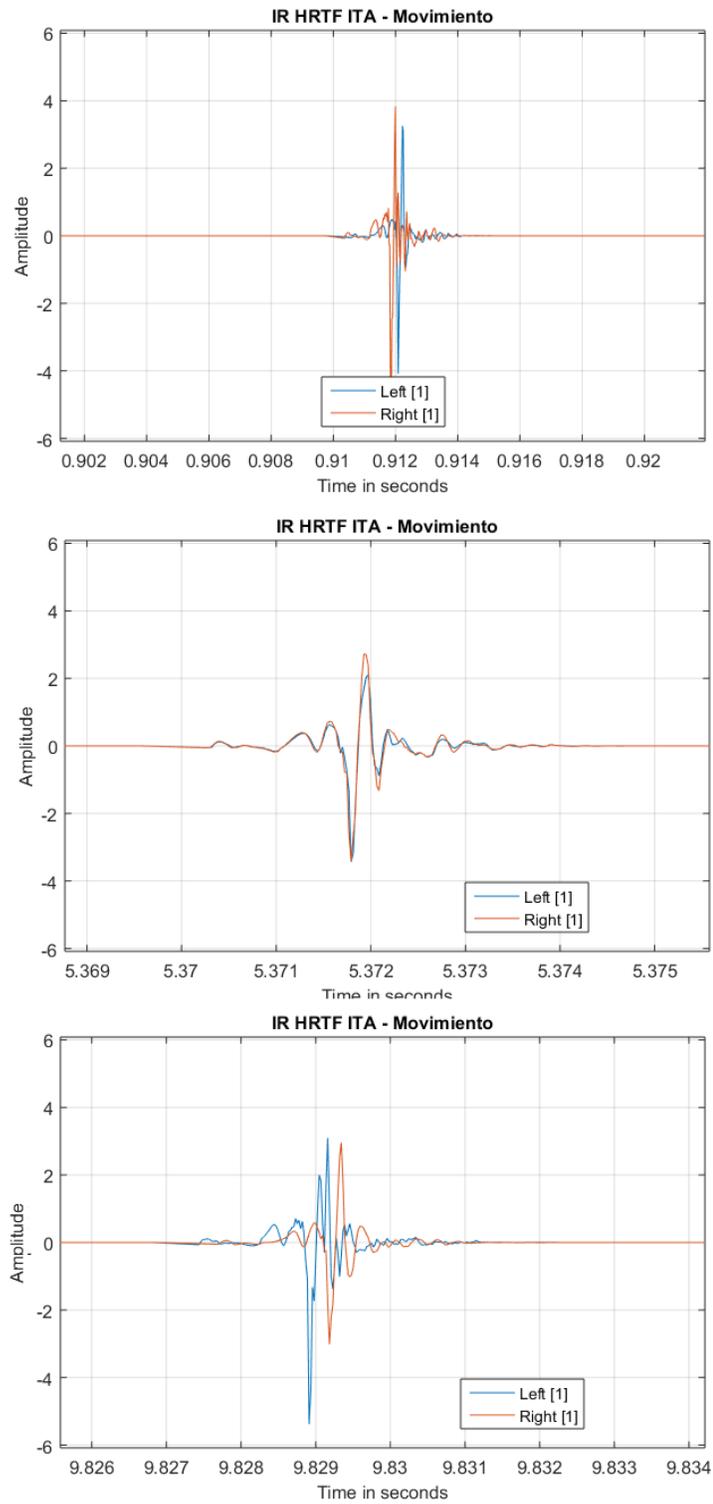


Figura 3-13: Respuesta de un impulso en movimiento (Primer, cuarto y último impulso) de HRTF ITA.

### 3.4 Simulación de audio 3D: SOFA

A través de Matlab, se puede cargar e inspeccionar cualquier base de datos SOFA, a través del comando “*SOFAload*”, el que se almacena dentro de una variable definida por el usuario. (Línea 5 A.4). En ese instante, se tendrá cargada HRTF en formato SOFA, quedando almacenada en una variable (“*hrtf*” en A.4). Análogamente, mostrará en pantalla la mayoría de los metadatos establecidos y asociados con la medición, receptores y fuentes de sonido. Alguna información adicional de la medición, tales como números de variaciones de grados que se utilizaron en la medición (Acimut, elevación, radio), puede consultarse mediante el comando “*SOFAinfo*”.

Análogo a esto, se pueden visualizar de forma gráfica algunas especificaciones de la medición a través del comando “*SOFAplotGeometry*”. En la figura 3-14a, se pueden visualizar todas las posiciones disponibles de fuentes, en función del receptor, su vista y posición de los canales de recepción. También se puede agregar una vista real de la simulación a ejecutar, visualizando en qué posición estaría la fuente sonora (Ver figura 3-14b).

Se debe dejar en claro que el formato “SOFA” se puede manejar de esta forma y mediante ITA toolbox, de la misma forma en la que se maneja un archivo “.*daff*”. Ésta sería una forma alternativa, cuya mayor ventaja sería el poder visualizar de forma gráfica el cómo están distribuidos los puntos de medición, tal y como se puede apreciar en la figura 3-14a.

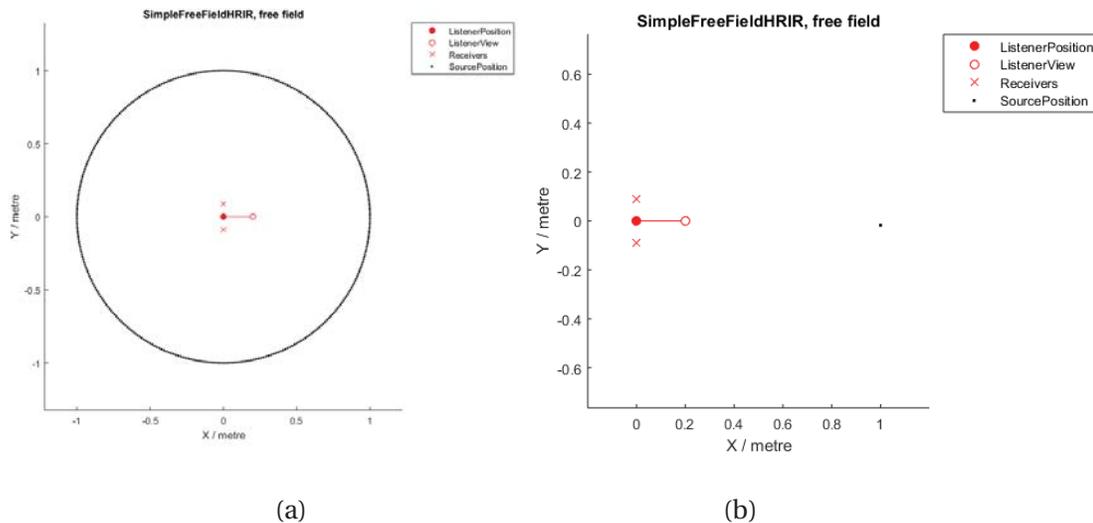


Figura 3-14: Gráfica de puntos de medición (a) y de simulación 3D en un punto frontal (b).

## 4 Mediciones de HRTF y HpTF

En el siguiente apartado se mostraran los resultados de mediciones de HRTF hechas a maniquís acústicos y mediciones de Headphone transfer function (HpTF).

### 4.1 Mediciones de HRTF

Se realizaron distintas mediciones de HRTF con diferentes variaciones en los grados acimutales. La primera medición se realizó cada  $20^\circ$  de variación en acimut, abarcando  $360^\circ$  en total, dando como resultado un total de 18 mediciones. En primera instancia, no fue necesario plantear una HRTF de “alta calidad”, como lo sería una medición realizadas con variaciones  $5^\circ$  o  $10^\circ$  en acimut. Esta medición ( $20^\circ$ ) se usó como forma de probar el método de medición, y a su vez para que su posterior procesamiento a través de Matlab, sean realizados de forma satisfactoria. En la figura 4-1 se puede apreciar una prueba preliminar que se hizo para probar el funcionamiento del sistema de mediciones.



Figura 4-1: Medición de HRTF en el laboratorio de acústica y vibraciones de la Escuela de Ingeniería Eléctrica (EIE).

Como se explicó anteriormente, a través de un script de Matlab (Ver Apéndice A.6), se asignaron las coordenadas correspondientes a cada medición, en el cual, análogamente se almacenaron en un archivo de la clase “itaHRTF” (ITA Toolbox). Finalmente se almacenaron estos datos en archivos de audio direccional, específicamente en los formatos ya estudiados: SOFA y DAFF.

#### 4.1.1 HRTF: Medición cada 20° en acimut

A través de SOFA y como se explicó en el capítulo 3, se procedió a visualizar de forma gráfica la posición del oyente, vista del oyente y posición de los canales receptores; esto en relación con las posiciones de la fuente sonora con la cual se realizó la medición. Con esto se verificó que las posiciones quedarán establecidas de forma correcta (Ver figura 4-2). Vale mencionar que al definir un archivo de formato tipo *.daff*, hasta el momento, se ha llegado a la conclusión de que se necesitan al menos 3 posiciones distintas en elevación, en los casos de mediciones cada 20° o 10° de variación en acimut. Debido a esto, se establecieron tres posiciones distintas en elevación, pero conformadas por el mismo grupo de mediciones, es decir, se establecieron falsas elevaciones. En primera instancia, esto no afectaría la simulación, ya que solo se está trabajando con variaciones acimutales sin elevación a la hora de realizar simulaciones.

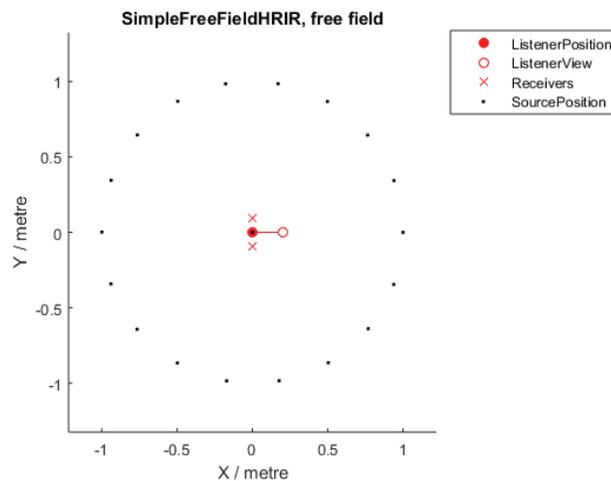


Figura 4-2: Posiciones establecidas para las posiciones de las fuentes que midieron la HRTF.

Una vez comprobada la correcta asignación de coordenadas a las mediciones, y a su vez, obtenido el archivo de audio direccional “*.daff*”, se realizaron pruebas con las simulaciones explicadas en el capítulo 3. Las cuales constan principalmente de generar fuentes sonoras virtuales fijas, y con movimiento alrededor del oyente. Para ambos casos se utilizó un impulso sonoro de corta duración.

El primer resultado se tuvo de la simulación de una fuente sonora virtual fija, estos resultados se presentan en la figura 4-3. En la parte superior izquierda de ésta, se presenta la respuesta impulsiva (RI) para una fuente sonora en la parte frontal del oyente (Figura 3-9), en la cual se puede apreciar cómo el pulso directo llega con niveles de amplitud similares en ambos canales y luego la amplitud va decreciendo de forma similar hacia las primeras reflexiones. Un efecto

distinto se puede apreciar en la RI del lado superior derecho, en el cual el pulso sonoro se ejecutó al lado izquierdo del oyente (lado opuesto a figura 3-6). Dado este caso, es posible apreciar cómo el pulso directo llega con una amplitud notoriamente mayor en el canal izquierdo (azul) en comparación al canal derecho (rojo), el cual además llega una fracción de tiempo después. En otras palabras, existe una diferencia de tiempo interaural (ITD) y diferencia de nivel interaural evidente (ILD). El ILD además es visible al observar las respuestas en frecuencia de este caso, donde se aprecia una diferencia de nivel, principalmente en altas frecuencias (efecto sombra, figura 1-5).

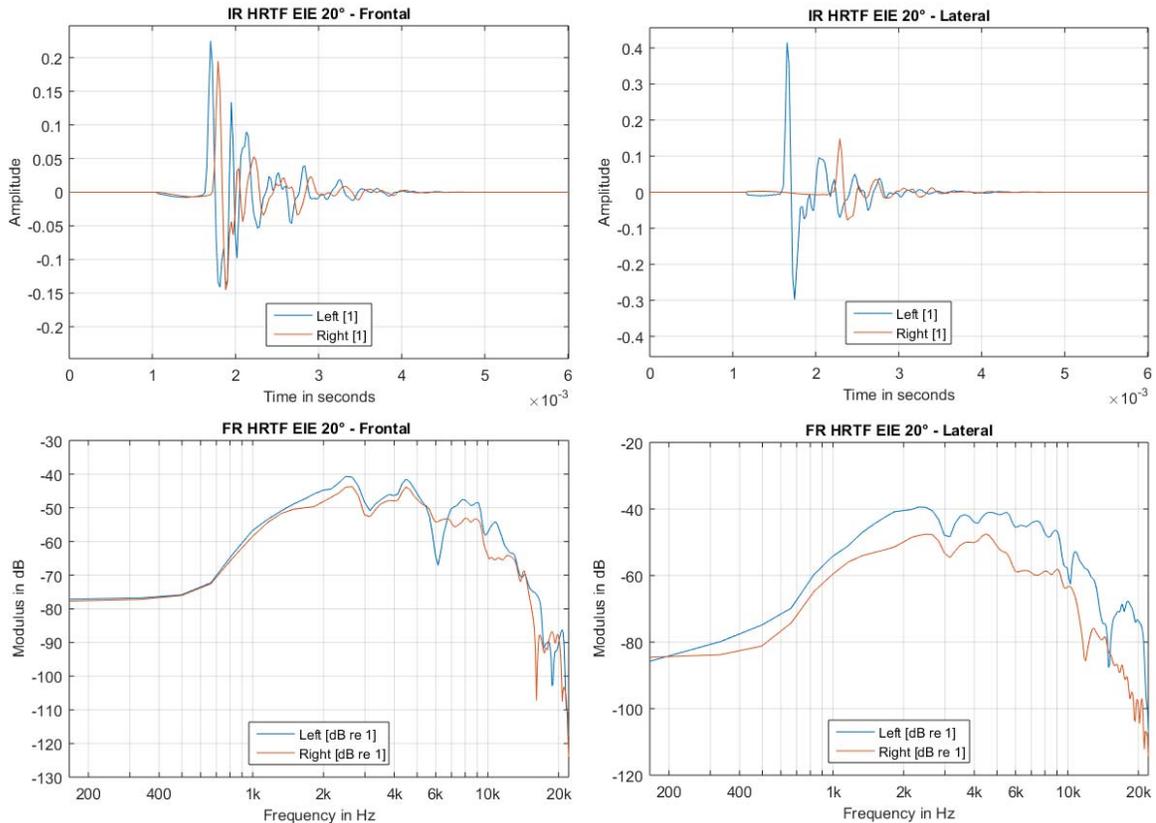


Figura 4-3: Respuestas impulso en el dominio del tiempo (arriba) y la frecuencia (abajo).

Para el caso en que el sonido se dio en movimiento alrededor del oyente, el pulso sonoro recorrió un trayecto semicircular de derecha a izquierda en sentido anti horario, específicamente con los parámetros establecidos en la Tabla 4-1 (estos parámetros se explicaron con detalle en el capítulo 3). Obteniendo la grabación de este resultado, se graficó en MATLAB obteniendo la respuesta impulsiva mostrada en la figura 4-5. En ésta, se puede apreciar cómo llega el pulso primero al canal derecho (rojo) mientras va disminuyendo en amplitud, así como análogamente va aumentando la magnitud del canal izquierdo (azul) a medida que este avanza en esa dirección. Análogo a esto se presenta un zoom de dos impulsos, para poder analizar en detalle la recepción de los pulsos (Figura 4-6), tiene coherencia con lo que se espera. Los pulsos que se presentan en la figura 4-6 corresponden a un pulso en la zona izquierda del oyente (Figura 4-6 izquierda) y en la zona frontal del oyente (Figura 4-6 derecha), dando un

comportamiento muy similar a los análisis realizados en la figura 4-3, por lo que no se entrara en detalles en cuanto a su interpretación.

Tabla 4-1: Datos del trayecto utilizado en la simulación.

Datos del trayecto		
circleR	3	radius of trajectory [m]
nlegs	400	number of equidistant trajectory legs
Tvel	13	time to pass nlegs points
phi_start	deg2rad(90)	start acimuth angle in [rad]
phi_end	deg2rad(270)	end acimuth angle in [rad]
theta	deg2rad(0)	zenith angle in [rad]

En la figura 4-4 se puede apreciar un esquema que puede facilitar la comprensión de la simulación que se llevara a cabo. El recorrido comienza desde  $\phi_{start}$  hasta  $\phi_{end}$  y sin elevación ( $\theta = 0^\circ$ ). El número de divisiones del recorrido se define como  $nlegs$ , que está representada por la línea segmentada, finalmente, la velocidad con que se recorre esta trayectoria está definida por  $Tvel$ .

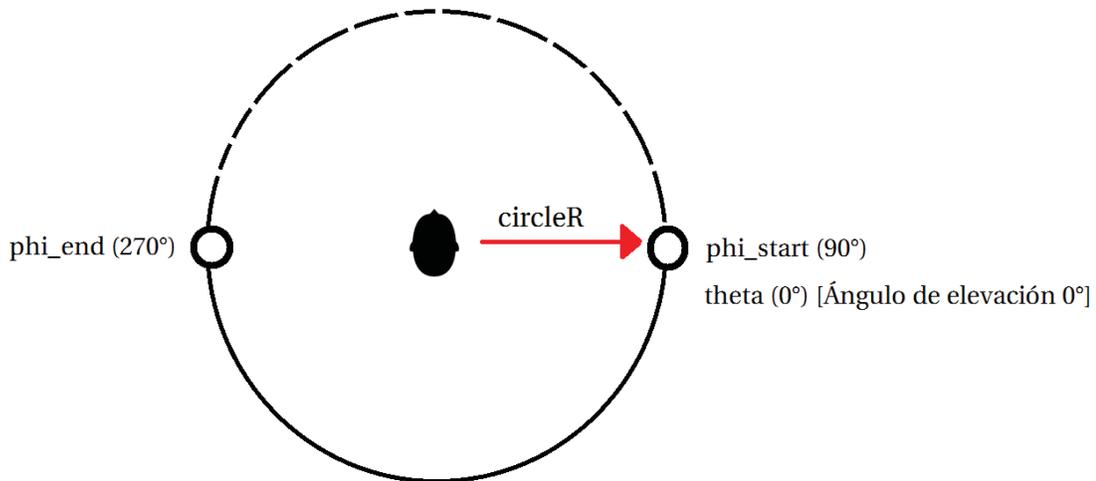


Figura 4-4: Representación gráfica del recorrido establecido en la tabla 4-1.

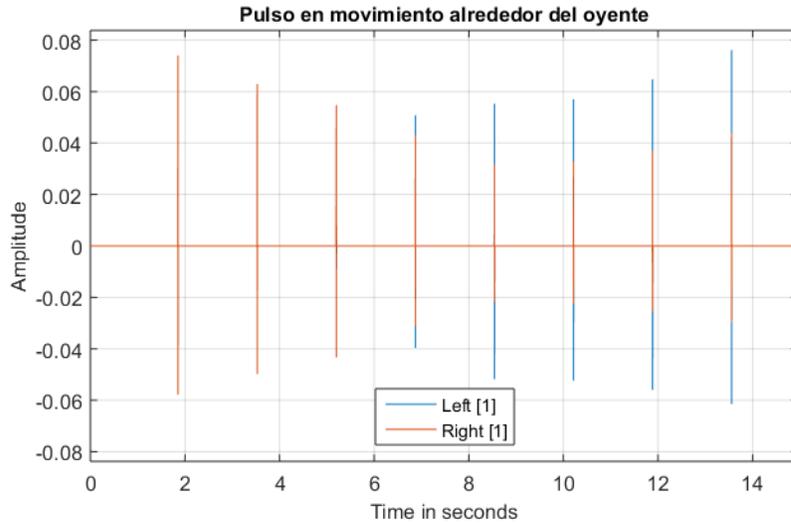


Figura 4-5: Grabación de pulso en movimiento alrededor (frente) al oyente.

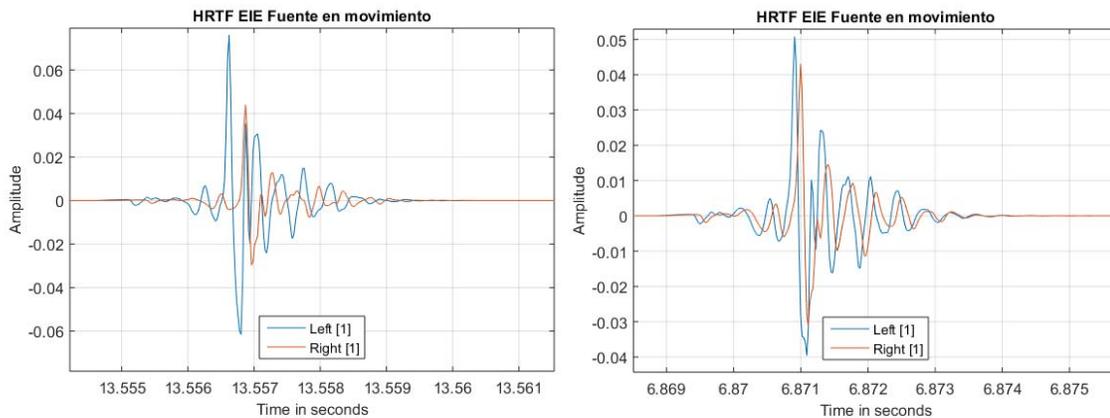


Figura 4-6: Detalle de IR de pulso en movimiento alrededor al oyente.

#### 4.1.2 ITA HRTF: Medición cada 10° en acimut

Esta medición se realizó cada 10° de variación en acimut y, al igual que la medición anterior, abarcando 360° en total, dando como resultado un total de 36 mediciones.

De igual forma que la medición anterior, se procedió a visualizar de forma gráfica la posición del oyente, vista del oyente, posición de los canales receptores (figura 4-7) a través de la forma explicada en el punto 3.4.

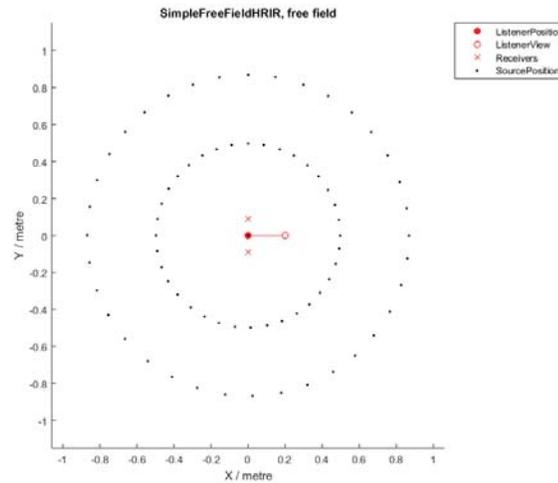


Figura 4-7: Posiciones establecidas para las posiciones de las fuentes que midieron la HRTF.

Los resultados del experimento simulado, se presentan en la figura 4-8. El análisis es similar al caso ya presentado anteriormente. En la parte superior de la figura 4-8, se puede observar una respuesta a un impulso (RI) en la zona frontal del oyente, en la cual se puede apreciar cómo el pulso sonoro llega con amplitud y tiempo similar para ambos canales y, a su vez, la amplitud va decreciendo hacia las primeras reflexiones. De igual forma, se puede apreciar una respuesta en frecuencia similar en nivel y forma, a excepción de pequeñas diferencias en altas frecuencias.

Un efecto distinto se puede apreciar en la RI del lado superior derecho, en el cual el pulso sonoro se ejecutó al lado izquierdo del oyente. Debido a esto, es posible apreciar cómo el pulso directo llega con una amplitud notoriamente mayor en el canal izquierdo (azul) en comparación al canal derecho (rojo), el que, además, llega una fracción de tiempo después, es decir, existe una diferencia de tiempo interaural y diferencia de nivel interaural evidente. Este último además es visible al observar las respuestas en frecuencia en la parte inferior de la figura 4-8, donde el efecto sombra se va haciendo cada vez más notorio a medida que aumenta la frecuencia.

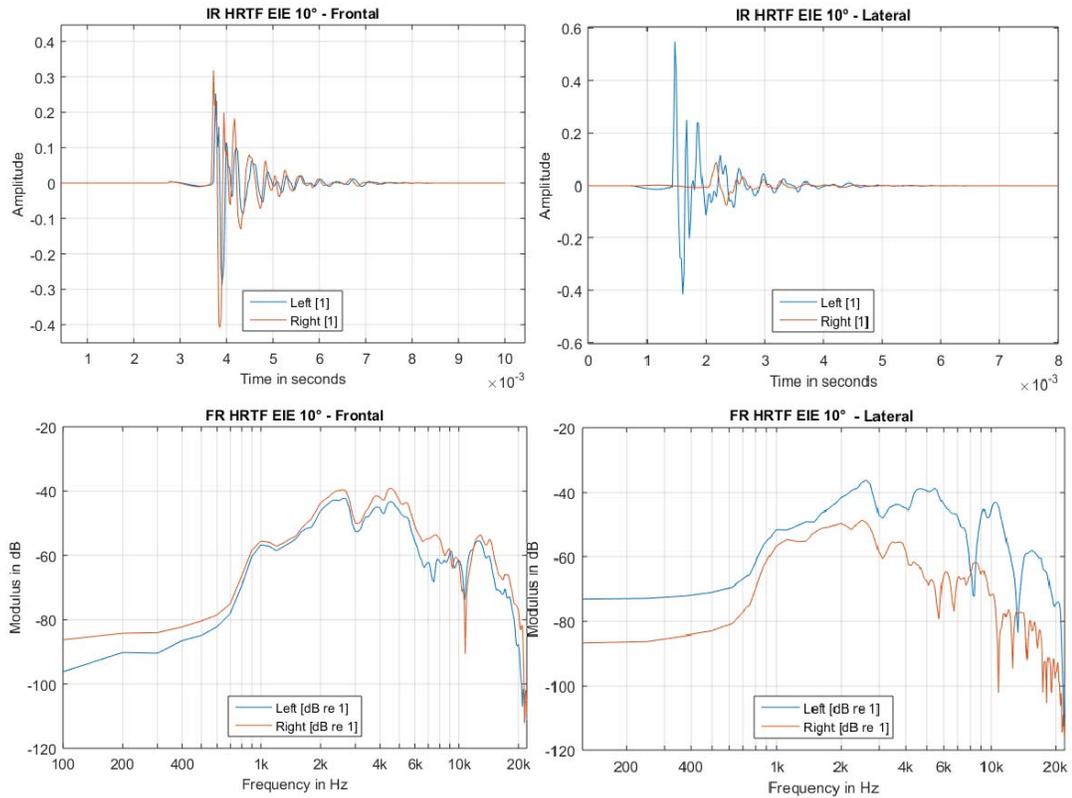


Figura 4-8: Respuestas impulso en el dominio del tiempo (arriba) y la frecuencia (abajo).

Para el caso de la fuente sonora en movimiento, se estableció el mismo recorrido que en la medición anterior (Tabla 4-1). Se graba el resultado de la respuesta a un impulso, dando como resultado la figura 4-9, donde se puede apreciar cómo llega el pulso primero al canal derecho (rojo) y como va disminuyendo en amplitud, así como análogamente va aumentando la magnitud del canal izquierdo a medida que éste avanza en esa dirección.

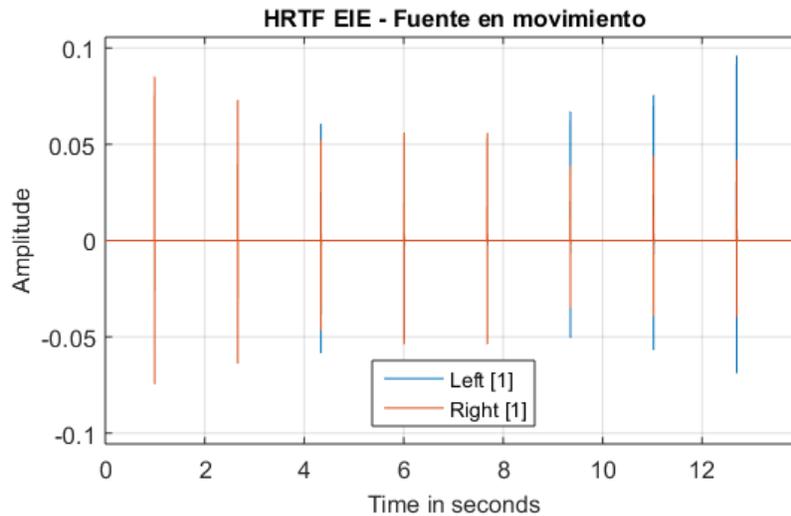


Figura 4-9: Grabación de pulso en movimiento alrededor (frente) al oyente.

Análogo a esto se presenta un zoom de dos impulsos, para poder analizar en detalle si la recepción de éstos tiene coherencia con lo que se espera (Figura 4-10). Los pulsos que se presentan en la figura 4-10 corresponden a un pulso en la zona izquierda del oyente (Figura 4-10 izquierda) y en la zona frontal del oyente (Figura 4-10 derecha), dando un comportamiento muy similar a los análisis realizados en la figura 4-3, por lo que no se entrara en detalles en cuanto a su interpretación.

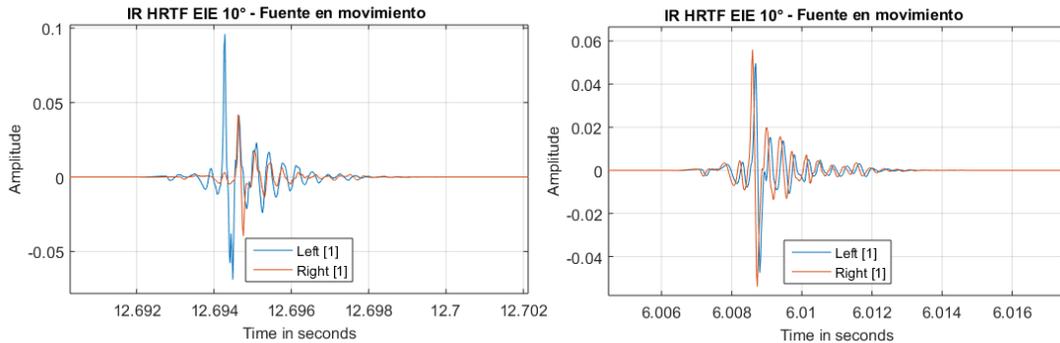


Figura 4-10: Detalle de IR e pulso en movimiento alrededor al oyente.

## 4.2 Medición de función de transferencia de audífonos

Como se mencionó anteriormente, con el objetivo de beneficiar la experiencia de inmersión en la simulación de audio 3D, se realizará una medición de la función de transferencia del audífono Sennheiser HD 600 (Figura 4-11). Debido a que de momento sólo se han obtenido mediciones de HRTF del ITA Dummy Head del Laboratorio de Acústica y Vibraciones (EIE, PUCV). Las mediciones de Headphone Transfer Function (HpTF) se realizaron en el mismo ITA dummy head. Para realizar esta medición se utilizó un pulso de barrido de frecuencia de 20[Hz] a 21 [kHz] y se realizaron aproximadamente 10 mediciones que se presentan en la figura 4-12. El procedimiento que se utilizó para realizar ésta medición fue el de sacar y volver a poner los audífonos cada vez que se realizaba un barrido de frecuencias, intentando variar la posición de la oreja dentro del audífono. Y es por esto que como resultado se obtuvieron mediciones distintas para cada experimento, sin embargo las atenuaciones o caídas en algunas frecuencias son similares y tienen coherencia en cuanto a una respuesta en frecuencia de una oreja. Finalmente para obtener una única medición, se obtuvo promedio, a través de MATLAB, y se trabajó con este resultado. En la figura 4-13 y 4-14 se pueden ver las respuestas en tiempo y frecuencia respectivamente.



Figura 4-11: Proceso de medición de HpTF.

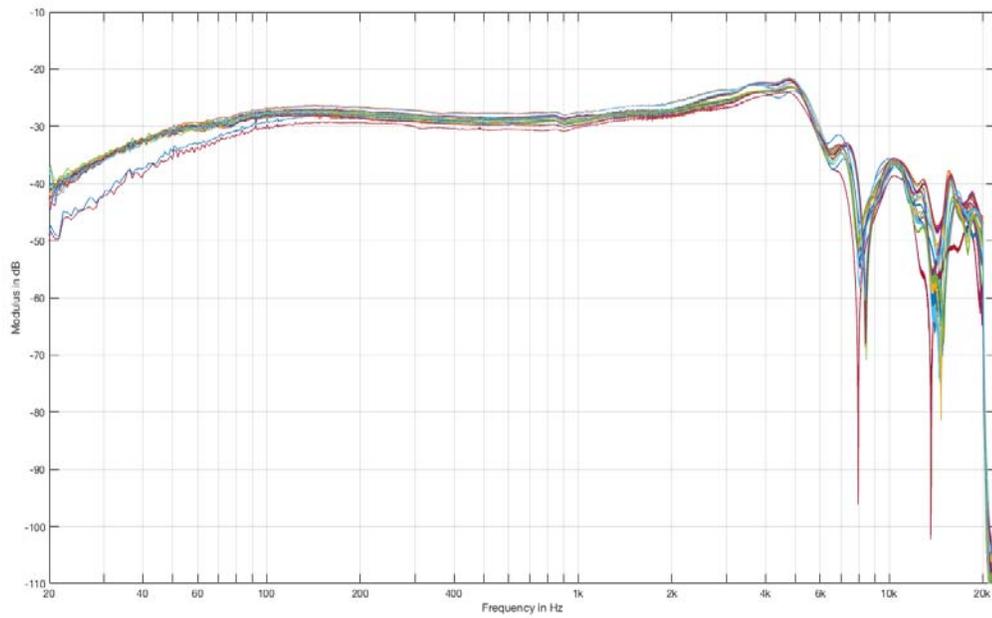


Figura 4-12: Respuesta en frecuencia de todas las mediciones realizadas.

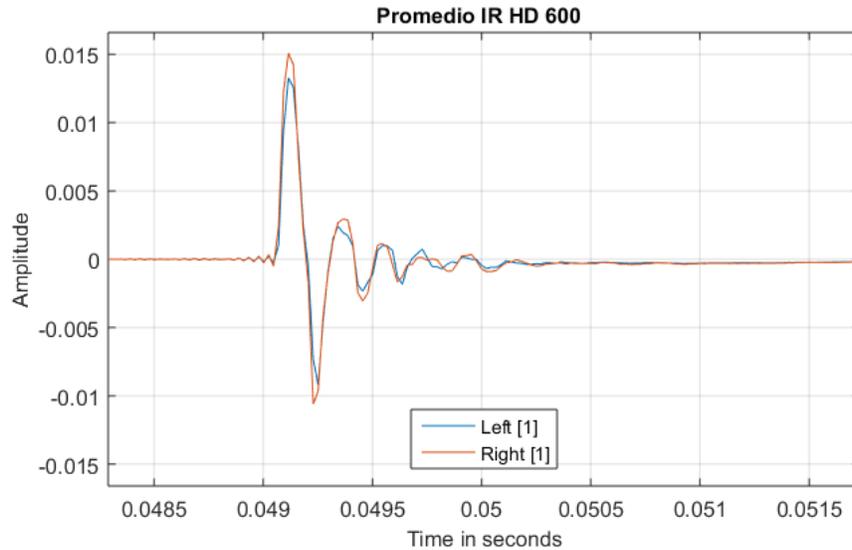


Figura 4-13: Promedio de la respuesta impulso de audífonos Sennheiser HD 600.

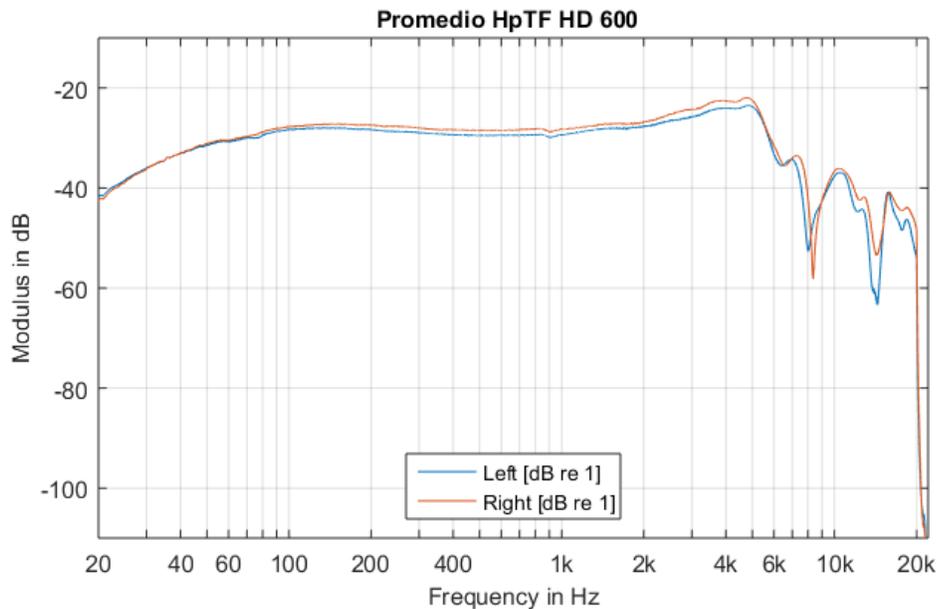


Figura 4-14: Promedio de la respuesta en frecuencia de audífonos Sennheiser HD 600.

Debido a su diseño de respaldo abierto y almohadilla para la oreja, estos auriculares tienen una entrega de bajos muy consistente a través de múltiples usuarios y reposiciones. Sin embargo, su entrega de agudos podría variar notablemente entre los usuarios y los asientos, dependiendo de las diferentes formas de las orejas y las preferencias de posicionamiento.

En la figura 4-15, se presentan mediciones sobre los audífonos Sennheiser HD 600 [32] y HD 605, como forma de referencia para poder analizar si las mediciones fueron representativamente correctas, no olvidando que existe mucha variabilidad en cuanto a formas y cuidados al realizar la medición.

En ambas figuras se pueden apreciar comportamientos similares. Como en la zona cercana al rango de 3[kHz] a 4[kHz] existe una leve amplificación, al igual que la zona de aproximada de 8[kHz] a 10[kHz]. En cambio, en las zonas cercana al rango de 5[kHz] a 6[kHz] existe atenuación, así como también existe en la zona de 10[kHz].

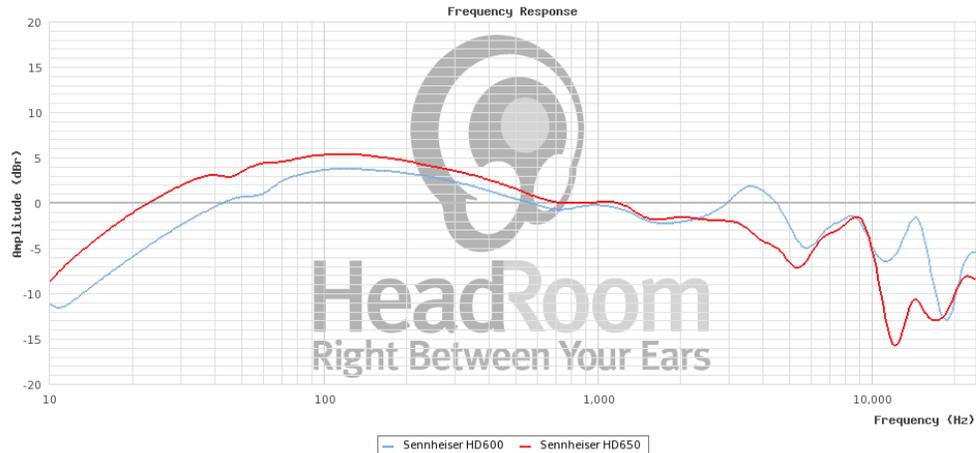


Figura 4-15: Respuesta en frecuencia de audífonos Sennheiser [32].

Se realizó un suavizado de las mediciones, atreves del comando “ita\_smooth\_frequency”. Con el fin de eliminar las fluctuaciones muy finas en la respuesta en frecuencia que no son perceptibles por el oído humano, al tiempo que conserva los detalles que son audibles para los humanos (Ver figura 4-16).

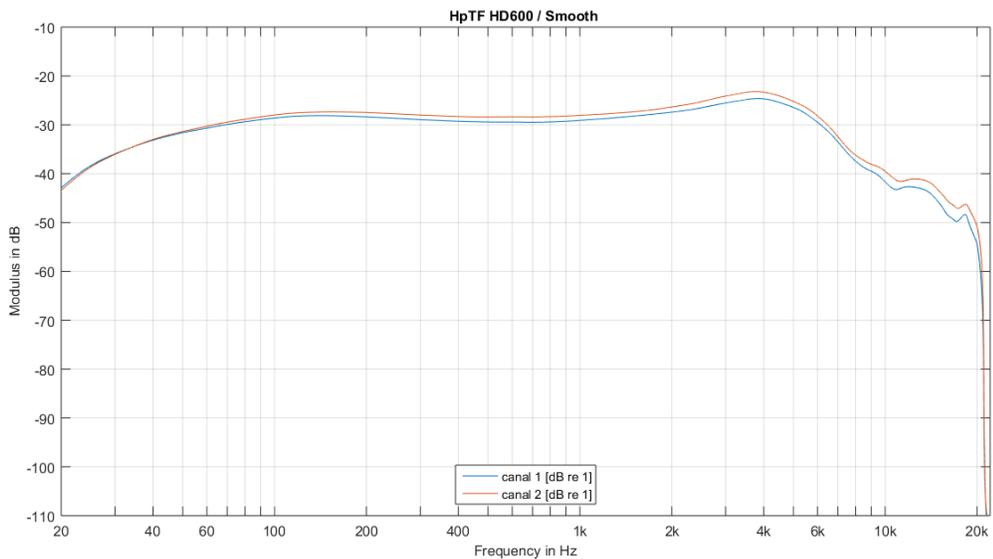


Figura 4-16: Respuesta en frecuencia suavizada de audífonos Sennheiser HD 600.

## 5 Implementación de Head-Tracking

En los siguientes apartados se procederá a describir algunas características y funciones relevantes a la hora de poder comprender el código que se usará para realizar el HeadTracking y que posteriormente se acoplará al sistema de audio 3D que se lleva a cabo mediante el Software Virtual Acoustic a través de Matlab.

### 5.1 Mtw Awinda – Xsens

El MTw Awinda es el rastreador inalámbrico de segunda generación de movimiento inercial de Xsens. El MTw permite en tiempo real aplicaciones cinemáticas 3D con múltiples rastreadoras de movimiento proporcionando una orientación muy precisa con una configuración de baja invasión.



Figura 5-1: The Xsens MTw Awinda hardware y MTw Sensor; Awinda dongle (USB) (tamaño 47[mm] x 30[mm] x 13[mm], peso 16 [g]) [33].

El MTw (Figura 5-1), es una IMMU (*inertial-magnetic measurement*) en miniatura. Para detectar el movimiento, el MTw contiene un sensor inercial cuyos componentes son: giroscopios de tasa 3D y acelerómetros 3D. Además, el MTw contiene magnetómetros 3D, un barómetro y un termómetro (usado para compensar la ganancia y sesgo dependencia de la temperatura de los otros elementos sensores). Combinando datos de giroscopios y acelerómetros (y magnetómetros), es posible registrar movimientos, sin una limitación en la captura movimiento y la necesidad de cámaras [34].

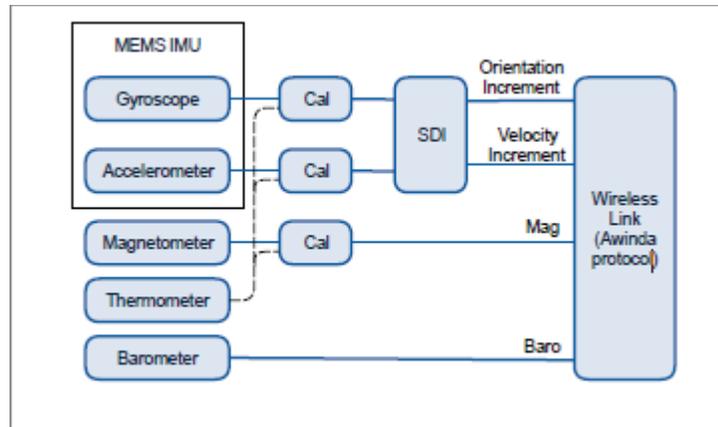


Figura 5-2: MTw Awinda signal processing architecture [34].

A bordo del sensor, el SDI se aplica a las calibradas lecturas del giroscopio y acelerómetro (Ver figura 5-2). El resultado de la SDI, junto con el magnetómetro calibrado y datos del barómetro, luego se transmite de forma inalámbrica utilizando el protocolo Awinda, al Maestro Awinda. El MTw se alimenta con una batería LiPo, con una duración de 6 horas. Está diseñado para ser robusto, fácil y cómodo en uso, con una colocación fácil en el cuerpo basada en flexibilidad correas de gancho y lazo.

Los datos de inercia de los giroscopios y acelerómetros son muestreados internamente a una frecuencia alta (1000 [Hz]) para garantizar la precisión en condiciones muy dinámicas, como la velocidad de movimientos, vibraciones e impactos. En base a estos datos, y utilizando un algoritmo de integración con abrazadera (SDI), el MTw calcula además los incrementos de velocidad y rotación, que luego se transmiten a través del enlace inalámbrico en un tasa de datos relativamente baja. [34]

Un dispositivo Xsens se conecta al sistema a través de un puerto de comunicación (representado por el cuadro "Controlador de comunicación serie" en la figura 5-3). La conexión de hardware subyacente para este puerto de comunicación puede ser un puerto serie o una conexión USB. A su vez, la obtención o recopilación de datos puede ser a través del software MT manager o a través de un script para Matlab. Este tipo de sensores viene diseñado para poder trabajar con varios sensores acoplados. Por ahora se limitara a trabajar con un solo dispositivo de medición, pero se dejara abierta la posibilidad de agregar más dispositivos.

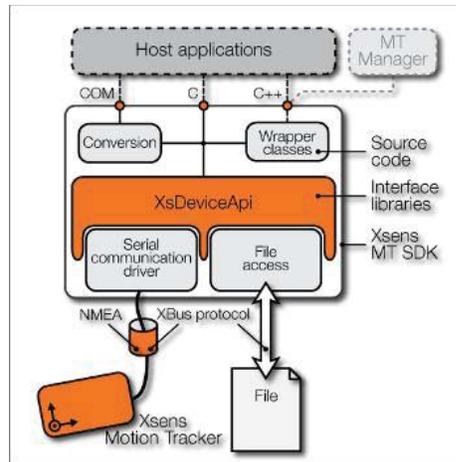


Figura 5-3: Configuración de desarrollo de aplicaciones

## 5.2 Estados de funcionamiento del sensor MTw

El MT tiene dos estados principales, el estado de configuración y medición. En el estado de configuración, varias configuraciones se pueden leer y escribir, y en el estado de medición, el MTw emitirá su mensaje de datos dependientes de la configuración actual. Una forma para entrar en estado de configuración o estado de medición es utilizar los mensajes *GoToConfig* o *GoToMeasurement*. El sensor trabaja con datos encriptados, que se requieren para poder procesar los datos mediante el software Xsens, para calcular valores de datos inerciales calibrados.

### Estado de configuración

El estado de configuración se utiliza para obtener y/o establecer varias configuraciones del MTw. La mayoría de los ajustes cambiarán la configuración que define la funcionalidad del dispositivo en el estado de medición. Dentro de las configuraciones se puede modificar, por ejemplo, la velocidad de transmisión de la comunicación, el período de muestreo, el modo de salida, la salida, entre otros. En el encendido, todos los ajustes se leen desde la memoria no volátil. Todos los ajustes se almacenan en un formato desarrollado por Xsens, conocido como eMTS (especificación ampliada de seguimiento de movimiento), junto con otros datos específicos del dispositivo como parámetros de calibración. Los ajustes modificados en el estado de configuración se almacenan inmediatamente en la memoria. Esta retendrá los últimos valores, incluso si el dispositivo está desconectado de la alimentación. Los cambios de configuración son inmediatos a excepción de la configuración de la velocidad en baudios, el cual requiere un reinicio.

## Estado de medición

En el estado de medición, el MTw enviará sus datos al host de una manera dependiendo de los ajustes definidos en el estado de configuración.

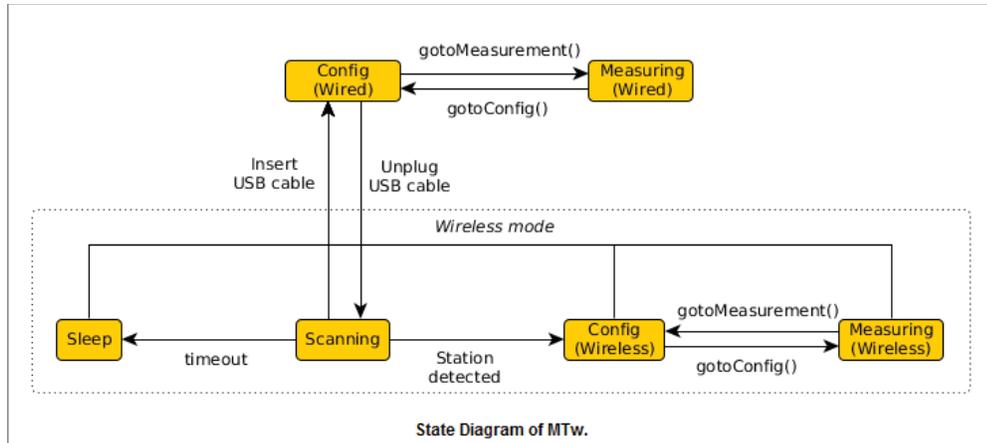


Figura 5-2: Configuración de desarrollo de aplicaciones

El SDK de dispositivos de Xsens implementa un sistema concurrente. Esto significa que las funciones XDA se implementan de forma segura para subprocesos. Para hacer que las funciones XDA sean seguras para subprocesos, la manipulación de datos se controla estrechamente y se debe tener cuidado de que las llamadas a funciones no sean bloqueadas.

## 5.3 Funcionamiento

A continuación se hará una breve descripción del modo de operación del sensor, esto será complementado con un diagrama de flujos presentado en la figura 5-3.

### Fase de inicio

1. Inicializar XDA (Xsense Device API): La aplicación comienza con la creación de un objeto *XsControl* utilizando la función *XsControl :: construct*.
2. Descubrir dispositivos: La aplicación detecta los dispositivos disponibles en el sistema mediante la función *XsScanner :: scanPorts*. Esta función devolverá todos los puertos de comunicación a los que están conectados los dispositivos Xsens.
3. Configurar dispositivos: Las acciones de configuración típicas son configurar el formato de datos de salida y la velocidad de datos a los valores deseados.
5. Preparación para el manejo de datos: Con el manejo de datos la aplicación se señala con un *XsDataPacket* tan pronto como llega. Para habilitar esto debe registrarse un objeto de devolución de llamada. La aplicación crea un objeto derivado de *XsCallback* que reemplaza la función *XsCallback :: onLiveDataAvailable*. El objeto se registra con XDA mediante la función *XsControl :: addCallbackHandler*.

6. Configurar Awinda Master: Establezca la tasa de actualización deseada (usando `XsDevice :: setUpdateRate`)
7. Habilitar la radio: Establecer el canal de radio y habilitar la radio. Los canales disponibles son de 11 a 26. Cuando la radio está habilitada, los MTw pueden conectarse de forma inalámbrica al maestro.
8. Cambia los dispositivos al modo de medición

Los dispositivos están configurados en modo de medición. La aplicación puede hacer esto por dispositivo principal con la función `XsDevice :: gotoMeasurement` o para todos los dispositivos simultáneamente usando `XsControl :: gotoMeasurement`

#### **Fase de procesamiento**

1. Manejar los datos entrantes: La función `onPostProcess` del objeto de devolución de llamada registrado en la fase de inicio se llamará siempre que haya nuevos datos disponibles. La función `XsDataPacket :: deviceId` se usa para determinar de qué dispositivo se recibe el paquete.
2. Comprobar la finalización: La aplicación verifica las condiciones requeridas para la finalización del proceso, estas son aplicaciones definidas. Si se cumplen las condiciones, se entra en la fase de salida.

#### **Fase de salida**

1. Cerrar dispositivos abiertos: Todos los dispositivos abiertos están cerrados. La aplicación utiliza la función `XsControl :: closePort` para cerrar todos los dispositivos abiertos. Cerrar el dispositivo cambiará automáticamente el estado del dispositivo del modo de medición al modo de configuración.
2. Limpiar XDA: La aplicación libera el objeto `XsControl` creado en el inicio utilizando la función `XsControl :: destruct`.

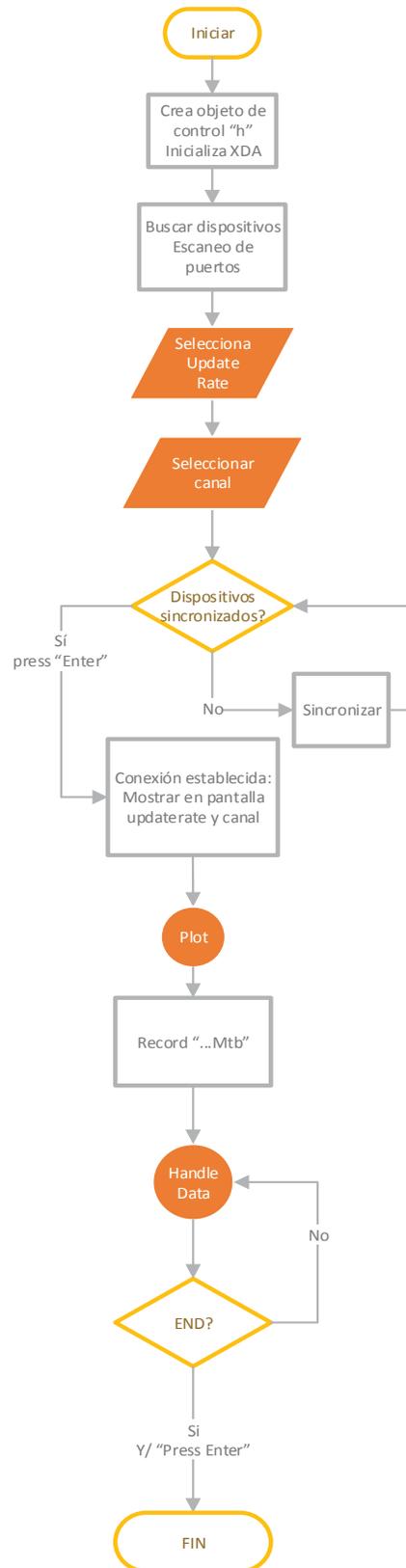


Figura 5-3: Algoritmo de funcionamiento del sensor Mtw.

## 5.4 Implementación

En este apartado se detallará el proceso de implementación del sensor dentro del software Virtual Acoustic que se opera a través de Matlab.

El funcionamiento de Virtual Acoustic ya se ha explicado con detalle en capítulos anteriores. Se complementará esto con un diagrama de bloques que explica a grandes rasgos las etapas de funcionamiento de Virtual Acoustic (Figura 5-4).

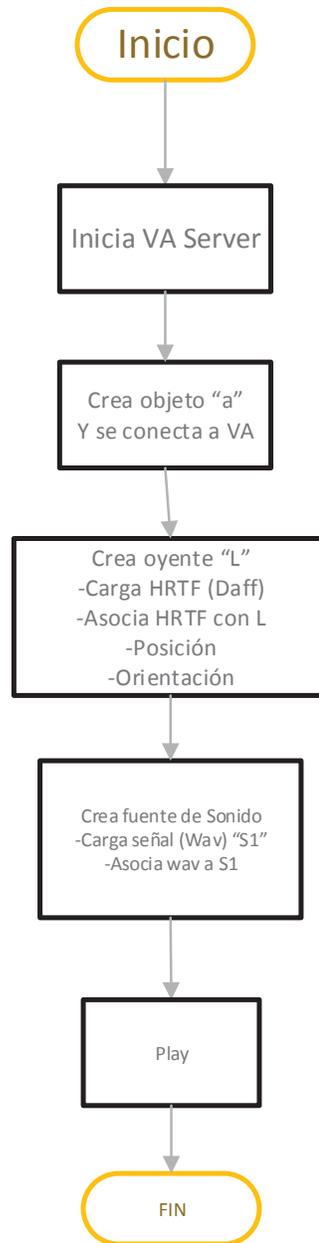
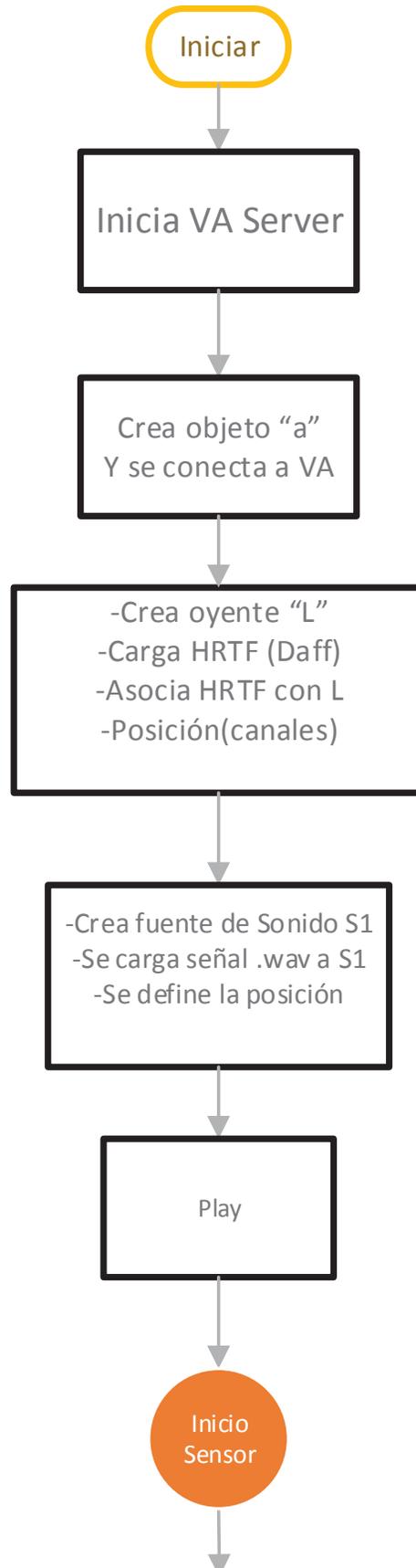


Figura 5-4: Diagrama de bloques del funcionamiento de VA.

Lo que se hizo fue mezclar ambos sistemas sin que estos se entorpecieran o se bloqueen entre ellos, debido a que se debe tener mucho cuidado en no bloquear las funciones llamadas durante el proceso de obtención de datos del sensor, ya que podría arrojar algún error de funcionamiento. Para poder facilitar la obtención de datos, se eliminaron todos los procesos que no serán realmente relevantes dentro del script de obtención de datos del sensor Mtw a través de Matlab. Algunos elementos importantes que fueron eliminados, fue todo lo relacionado con la visualización gráfica de los datos obtenidos en tiempo real. Así, solo se trabajaría con la obtención de datos. Se presenta un algoritmo en la figura 5-5 con la fusión de ambos sistemas.

Se partirá iniciando los elementos necesarios de Virtual Acoustic, por lo que la simulación acústica comenzará antes de el acoplamiento del sensor en el sistema, es decir, se tendrá la simulación acústica corriendo unos segundos, mientras el sensor se sincroniza y se pone en condiciones para poder iniciar el headtracking. Una vez ocurrido esto llega la etapa de llamada de la función “handle data”, la cual tiene el rol de llevar un orden de todos los paquetes de datos que se reciben del sensor, puedan llegar de forma correcta y ordenada para ser interpretada. Es en esta etapa donde se van recibiendo los nuevos paquetes pertenecientes a los datos de orientación, en la que se agregó la etapa donde se establece la orientación del oyente.



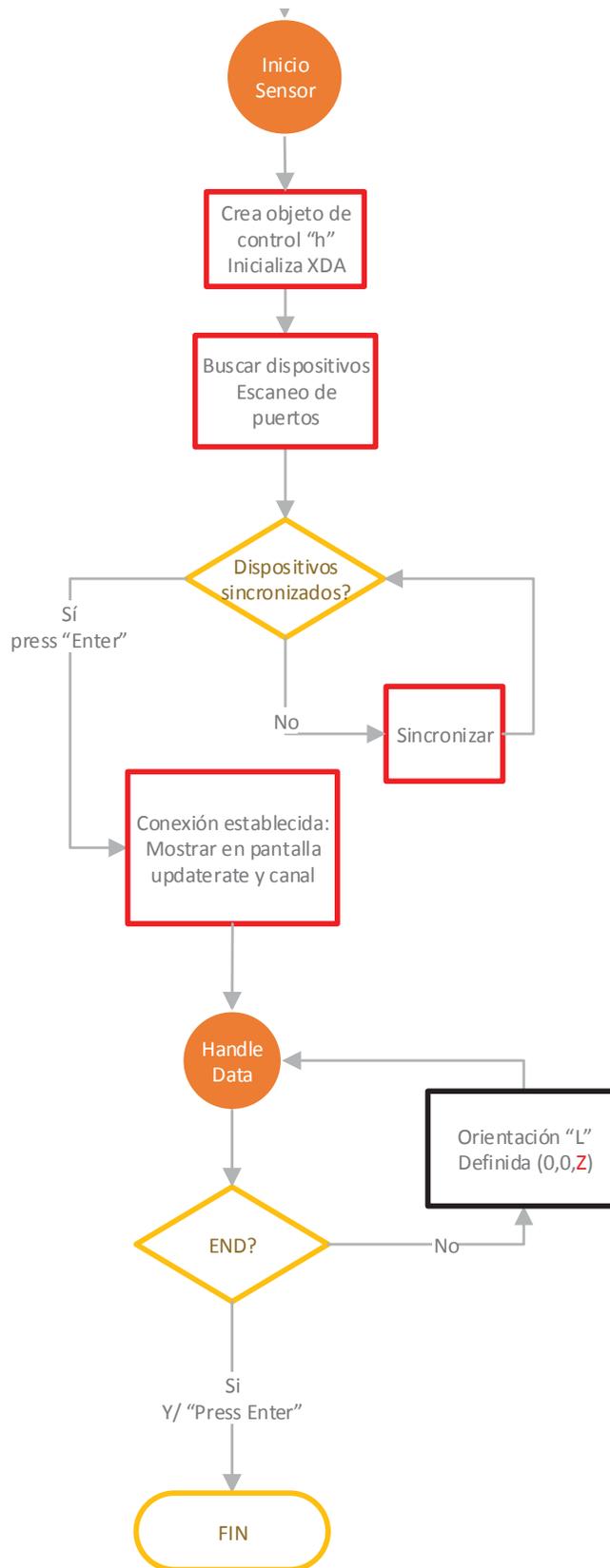


Figura 5-5: Diagrama final del funcionamiento del sensor mtw junto a VA.

## 5.5 Resultados

Se realizaron pruebas para poder analizar el correcto funcionamiento del sensor de movimiento Mtw. Se partió haciendo una prueba simple de rotación del oyente (sensor) dentro de una simulación de audio 3D con fuente fija en la escena, donde la fuente permaneció siempre delante del oyente (Ver figura 5-6), además para esta fuente sonora se utilizó un impulso sonoro de corta duración. Análogo a esto, el oyente (sensor) debía girar hasta completar una vuelta completa hasta llegar a la posición original. Se espera poder apreciar de forma gráfica las variaciones de los ITD e ILD según corresponda el caso.

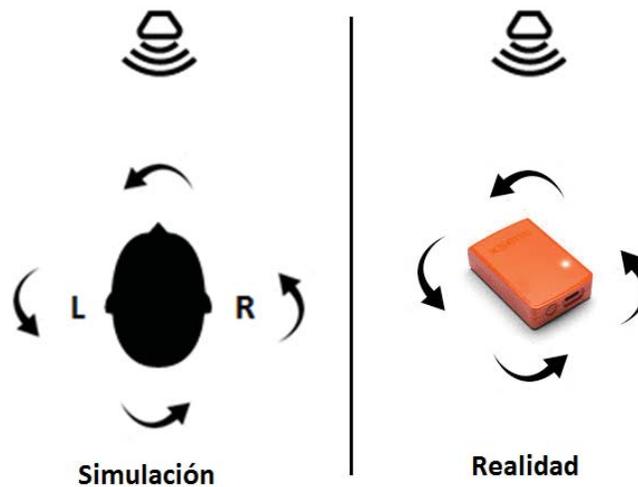


Figura 5-6: Diagrama de experimento para prueba de sensor.

Los resultados presentados (figura 5-7), reflejan de forma satisfactoria lo que se esperaba en un principio. En ésta se puede apreciar cómo existe una etapa inicial que va desde el inicio hasta el segundo 12 aproximadamente, donde el sensor aún no ha sido activado dentro de la simulación. Una vez configurada las opciones del sensor y a su vez esperando el tiempo de sincronización necesario, se activa el sensor. Se empieza con el movimiento del sensor (oyente) hacia la izquierda, dejando más susceptible al pulso directo al canal derecho (rojo) y, contrariamente, dejando más alejado del pulso directo al canal izquierdo (azul). Luego hay un momento donde los niveles se igualan nuevamente, es decir el oyente ahora se encuentra dándole la espalda a la fuente sonora. Seguido de esto los roles se invertirían, dejando ahora más susceptible al pulso directo al canal izquierdo (azul), contraponiéndose el canal rojo de la misma forma que se explicó. Una vez finalizado el experimento, es decir, recorriendo 360° de rotación en el plano Z (yaw), se procede a desactivar el sensor, dando un pequeño lapso de finalización similar al del inicio y se termina con la simulación. Análogamente se presentan los resultados con el canal izquierdo (azul), desplazado una fracción de segundo, para que así, se pueda apreciar mejor la diferencia de niveles que existe en cada intervalo de tiempo (Ver figura 5-8).

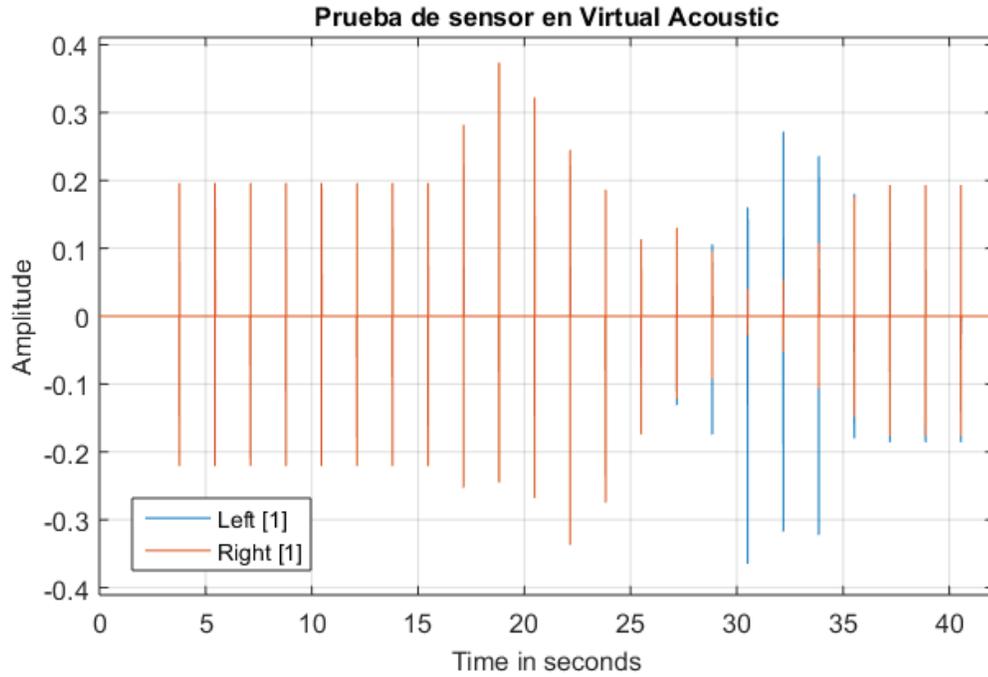


Figura 5-7: Resultado simulación audio 3D con Headtracking.

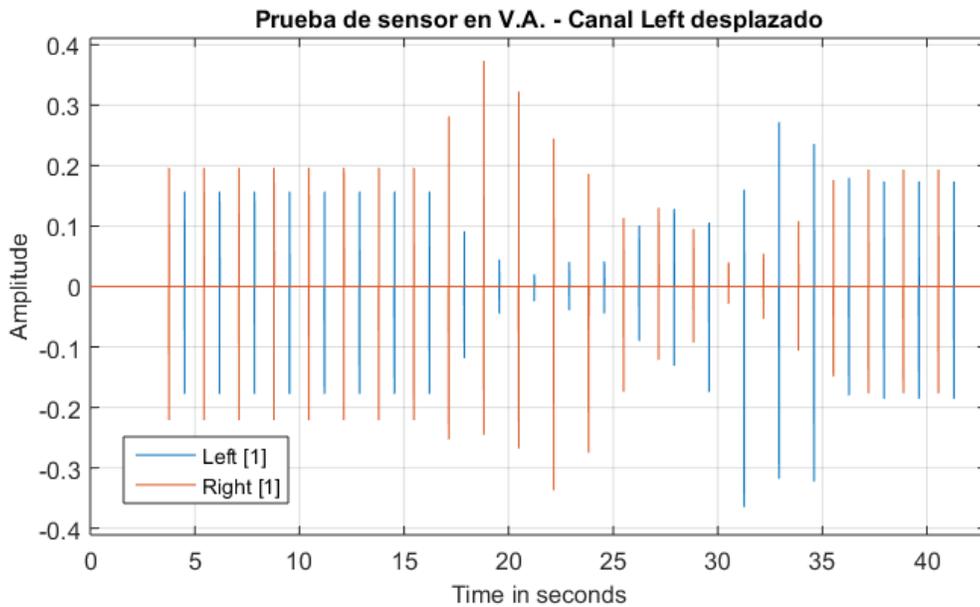


Figura 5-8: Resultado simulación audio 3D con Headtracking (canal izquierdo desplazado en el tiempo).

## Conclusiones

En el comienzo de este trabajo, fue posible apreciar el complejo sistema que implica recrear una sensación de escucha tridimensional. Basándose en un estudio del comportamiento de la audición humana y sus diferentes técnicas de localización, donde la teoría dúplex fue fundamental a la hora de poder analizar los resultados de las simulaciones como de las mediciones de HRTF. Por otro lado se investigó sobre sistemas de audio que intentan replicar el sonido 3D, donde destacaron la reconstrucción binaural, específicamente por convolución.

Una vez comprendido el fundamento y el estado actual de experimentos para generación de sonidos 3D, se estudiaron algunas herramientas específicas para poder realizar una escena virtual de sonido 3D. Una de las herramientas estudiadas más importantes dentro del proyecto, sin duda, fue *Virtual Acoustic*, el cual fue elegido como el actor principal dentro de la auralización y configuración de la escena acústica virtual, sin dejar de lado MATLAB que es el software escogido para trabajar con VA. Por otro lado, MATLAB y su herramienta “ITA Toolbox” permitió manipular las simulaciones y mediciones, así como su post procesamiento y comparaciones con otras bases de datos.

Debido a la gran cantidad de formatos nativos de diferentes laboratorios alrededor del mundo, se estudiaron algunos formatos para poder recopilar datos de audio direccional (SOFA y Daff). De estos formatos, Virtual Acoustic solo puede trabajar con el formato Daff, pero VA también sugiere trabajar con SOFA, que sirvió para poder comprobar la correcta configuración de los datos de medición. Análogo a esto, a través de “ITA Toolbox”, se pudieron realizar las conversiones de formato necesarias para daff, SOFA y wav, así como manipulación de señales acústicas en general.

Se logró implementar diferentes tipos de simulación 3D. Donde las simulaciones de fuentes fijas y en movimiento alrededor del oyente, fueron muy relevantes a la hora de realizar experimentos que permitieron estudiar la calidad de las mediciones de HRTF medidas por el laboratorio de acústica y vibraciones de la EIE. Donde se logró ver de forma gráfica las variaciones de ITD e ILD, para diferentes casos según la ubicación del par fuente-receptor.

Finalmente se logró implementar un sensor de movimiento para poder realizar *HeadTracking* dentro de una simulación con fuente fija en el espacio, esto dará más de libertad al movimiento de cabeza durante la simulación. Ya que el sonido resultante de la simulación, se ira

modificando según la orientación de la cabeza del oyente, dando como resultado una simulación mucho más consistente.

En cuanto a los resultados que se obtuvieron, se quedó bastante conforme en cuanto al avance establecido durante el proceso de implementación, logrando superar cada vez más la calidad de las mediciones, así como su post procesamiento en los formatos de audio direccional DAFF y SOFA. Logrando finalmente tener un archivo de HRTF del maniquí acústico *ITA dummy head*, la cual se realizó cada 20° grados de variación en acimut y fue llevada a cabo íntegramente por el equipo de trabajo perteneciente al FONDECYT #11150388. Ésta medición se puso a prueba mediante las simulaciones que se explicaron anteriormente, en la cual se logró apreciar de forma gráfica las distintas variaciones de ITD e ILD para distintas posiciones de fuentes virtuales, así como fuentes en movimiento.

Para finalizar, hay que mencionar que no se cumplieron todos los objetivos planteados inicialmente. Existieron algunos problemas que dificultaron el avance y el poder realizar completamente comparaciones de HRTF individualizadas. Los procesos de obtención de maniqués acústicos a través de fotogrametría no presentaron los resultados esperados con rapidez, lo cual retrasó todo el proceso de obtención de maniqués individualizados, el cual a su vez retrasó el proceso de medición del maniquí. Otro inconveniente que se tuvo, fue con el sistema que permitía girar el maniquí acústico en el plano acimutal al momento de realizar mediciones. A su vez, la plataforma que iba a permitir realizar mediciones con elevación, no se pudo armar debido a que las piezas de madera con la cual fue fabricada ya no lograban ser encajadas según el diseño original, dejándola, de momento, inservible. Aun así se avanzó bastante en la medición de HRTF, solucionando muchos problemas que fueron saliendo en el camino, de esta forma, se sentaron buenas bases para seguir midiendo HRTF en el futuro.

Dentro de las proyecciones y trabajos futuros, se deben realizar unas mediciones de HRTF con mayor resolución, es decir cada 5° o 10° grados en acimut y a su vez complementarla con mediciones en elevación. Es muy importante realizar, experimentos de localización con y sin *Headtracking*, para poder estudiar su relevancia frente a un experimento con movimiento de cabeza o posición fija. Por otro lado, se deja abierta la posibilidad de poder plantear una futura comparación de grabación binaural y síntesis binaural a través de convolución.

# Bibliografía

- [1] M. Möser, *Ingeniería Acústica: Teoría y Aplicaciones*, 2009.
- [2] F. C. Tommasini, M. Araneda y O. A. Ramos, «Auralización en tiempo real: Implementación del modelo del oyente,» *2o Congreso Internacional de Acústica UNTREF*, 2010.
- [3] O. A. Ramos, G. Calvo y F. C. Tommasini, «MODELO ACÚSTICO DE CABEZA Y TORSO MEDIANTE ANÁLISIS DE COMPONENTES PRINCIPALES,» *Centro de Investigación y Transferencia en Acústica, CINTRA. UA-CONICET UTN, Regional Córdoba.*, 2007.
- [4] E. J. Angel, V. R. Algazi y a. R. O. Duda, «On the design of canonical sound localization environments,» *Interface laboratory, CIPIC, University of California, USA*.
- [5] Oculus, «Developer Oculus,» [En línea]. Available: <https://developer.oculus.com/documentation/audiosdk/latest/concepts/audio-intro-localization/>.
- [6] Á. L. Xaus, «Sonido 3D, La percepción acústica egocéntrica y principios de sistemas espaciales,» Barcelona, 2009.
- [7] D. Maggiolo, «Escuela Universitaria de Música,» Universidad de la República, 2003. [En línea]. Available: <https://www.eumus.edu.uy/>.
- [8] P. M. A. Hofman, J. G. V. R. and y J. V. Opstal, «Relearning sound localization with new ears,» 1998.
- [9] E. A. Macphersonb y J. C. Middlebrook, «Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,» *Kresge Hearing Research Institute; University of Michigan*, 2002.
- [10] J. Merimaa, «Analysis, Synthesis, and Perception of Spatial Sound - Binaural Localization

- Modeling and Multichannel Loudspeaker Reproduction,» *Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing*, 2006.
- [11] T. Francart, «Perception of Binaural Localization Cues with Combined Electric and Acoustic,» Katholieke Universiteit Leuven, 2007.
- [12] M. Crespo, «hisponic,» 2017. [En línea]. Available: <https://www.hisponic.com/tutoriales/como-ubicamos-sonidos-espacio-binauralidad-teoria-duplex/43279>.
- [13] «Waves,» 2016. [En línea]. Available: <https://www.waves.com/3d-audio-on-headphones-how-does-it-work>.
- [14] I. Rossell y R. Barti, «Design and implementation of a 3D sound system based on the HRTF,» *Department d'Acústica, Enginyeria La Salle*.
- [15] C. I. Cheng y D. G. H. Wakefield, «Introduction to Head-Related Transfer Functions (HRTF'S): Representations of HRTF'S in time, frequency, and space».
- [16] E. Choueiri, «princeton,» Princeton University, [En línea]. Available: <https://www.princeton.edu/3D3A/HRTFMeasurements.html>.
- [17] K. H. SHI y Y. PARK, «Enhanced Vertical Perception through Head-Related Impulse Response Customization Based on Pinna Response Tuning,» *IEICE TRANS FUNDAMENTALS*, vol. VOL.E91-A, 2008.
- [18] M. Benjamin, «vivavangelder,» 2014. [En línea]. Available: <http://vivavangelder.blogspot.com/2014/02/el-paso-del-sonido-mono-al-estereo-en.html>.
- [19] M. O. Nave, «hyperphysics,» Georgia State University, 2000. [En línea]. Available: <http://hyperphysics.phy-astr.gsu.edu/hbasees/Audio/audsig.html>.
- [20] SDAC , «SDAC (System Dynamics and Applied Control),» NoViC (Noise and Vibration control Center) , [En línea]. Available: <http://sdac.kaist.ac.kr/project/index.php?mode=completed&act=SFRT>.
- [21] W. G. Gardner, «3D Audio using Loudspeakers,» *Massachusetts Institute of Technology*, 1997.
- [22] ITA, «Virtual Acoustics Institute of Technical Acoustics,» RWTH Aachen University, [En línea]. Available: <http://virtualacoustics.org>.
- [23] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger y M. Noisternig, «Spatially Oriented Format for

- Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions,» *Audio Engineering Society*, 2013.
- [24] ITA, «OpenDaff,» Institute of Technical Acoustics, Aachen, [En línea]. Available: <http://www.opendaff.org/>.
- [25] F. Wefers, «OpenDAFF A free, open-source software package for directional audio data,» *Institut für Technische Akustik (ITA), RWTH Aachen, Daga (Berlin)*, 2010.
- [26] AES69-2015, «SOFA Conventions,» [En línea]. Available: <https://www.sofaconventions.org/>.
- [27] i. Audio Engineering Society, «AES standard for file exchange-Spatial acoustic data file format,» *Audio Standards*.
- [28] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger y M. Noisternig, «Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions,» *Audio Engineering Society*, 2013.
- [29] P. Dietrich, B. Masiero y R. Scharrer, «Application of the MATLAB ITA-Toolbox: Laboratory Course on Cross-talk Cancellation,» *Institute of Technical Acoustics, RWTH Aachen University*, 2011.
- [30] P. Dietrich, M. Guski, M. Pollow y M. Muller-Trapet, «ITA-Toolbox - An Open Source MATLAB Toolbox for Acousticians,» *Institute of Technical Acoustics, RWTH Aachen University*, 2012.
- [31] G. G. Nyambuya, «Bipolar outflows as a repulsive gravitational phenomenon - Azimuthally Symmetric Theory of Gravitation (II),» *Research in Astron and Astrophys*, 2010.
- [32] headphone. [En línea]. Available: <https://www.headphone.com/>.
- [33] xsens, «MTw Awinda Wireless Motion Tracker,» [En línea]. Available: <https://www.xsens.com/products/mtw-awinda/>.
- [34] M. S. a. G. B. Monique Paulich, «Xsens MTw Awinda: Miniature Wireless Inertial Motion Tracker for Highly Accurate 3D Kinematic Applications,» *XSENS TECHNOLOGIES*.

# A Códigos de Matlab

En este apéndice se muestran los códigos relevantes que se utilizaron para realiza las simulaciones de audio tridimensional utilizando MATLAB y Virtual Acoustic.

## A.1 Proceso para cargar y visualizar datos de un archivo DAFF y SOFA

Como ya se ha estudiado anteriormente, estos archivos almacenan datos de audio direccionales, es decir, respuestas impulsos de cabezas de pruebas (dummy head) asociadas a distintas posiciones. Se procederá a explicar de forma concisa como cargar estos archivos a Matlab y como revisar sus mediciones.

Para cargar un archivo daff o sofá, se usa la función “itaHRTF”, especificando el formato del archivo y el nombre y/o ruta, dependiendo de si el archivo se encuentra o no en la misma carpeta en la que se está trabajando en Matlab. Si se está trabajando en la misma carpeta donde se encuentra el archivo, bastaría solo con el nombre.

Finalmente el código quedaría de la siguiente forma:

```
hrtf_sofa = itaHRTF('SOFA', 'TUBerlin_QU_KEMAR_anechoic_radius_1m.sofa');
```

```
hrtf_daff = itaHRTF('DAFF', 'ITA_Artificial_Head_5x5_44kHz_128.v17.ir.daff');
```

Quedando almacenado en las variables hrtf\_sofa y hrtf\_daff respectivamente.

Ahora para poder acceder a las mediciones que guardan estos archivos, se puede acceder a todas las mediciones juntas con un simple “.plot\_time” para revisar todas las respuestas impulso o “plot\_freq” para las respuesta en frecuencia, acompañado del nombre de la variable a revisar antes del punto, por ejemplo: hrtf\_daff.plot\_time . (Figura A-1 y A-2)

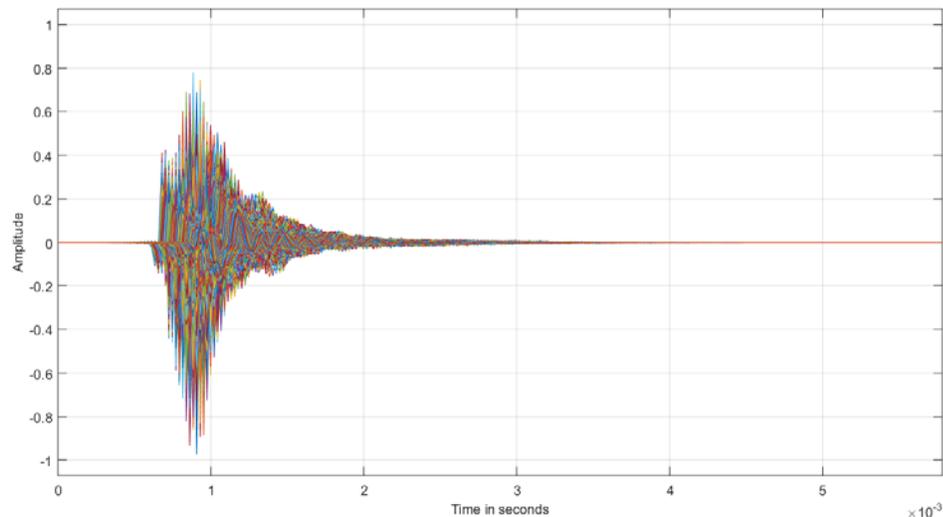


Figura A-1: Todas las respuestas a impulsos graficadas, correspondientes a la HRTF  
“TUBerlin\_QU\_KEMAR”

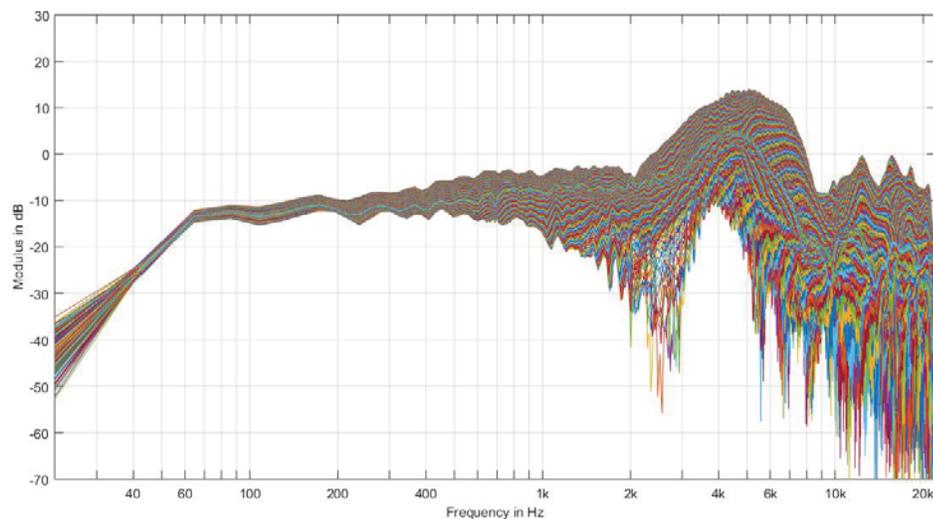


Figura A-2: Todas las respuestas en frecuencia graficadas, correspondientes a la HRTF  
“TUBerlin\_QU\_KEMAR”

Debido a la gran cantidad de mediciones que se encuentran almacenadas en estos archivos, no es muy conveniente revisarlas de esta forma. Existe una forma más específica de revisar una medición en una coordenada definida por el usuario. Para esto se debe especificar la coordenada a revisar mediante el comando “itaCoordinates”, especificando: radio de la medición (pueden existir mediciones hechas a más de 1 radio de distancia), elevación (theta) y acimuth (phi) y especificando además en que sistemas de coordenadas se está consultando, en este caso se está consultando en coordenadas esféricas por tanto la coordenada debería ir acompañada de “,’sph””. El comando quedaría de la siguiente forma.

```
coordF = itaCoordinates([1 pi/2 0], 'sph');
```

Ya se tiene la coordenada, ahora se debe consultar la coordenada del archivo, mediante el comando.

```
HRTF_find = hrtf_sofa.findnearestHRTF(coordF);
```

Y finalmente, ya se podría graficar el resultado almacenado en la variable "HRTF\_find" mediante el comando "plot\_freq" o "plot\_time" (Figuras A-3 y A-4)

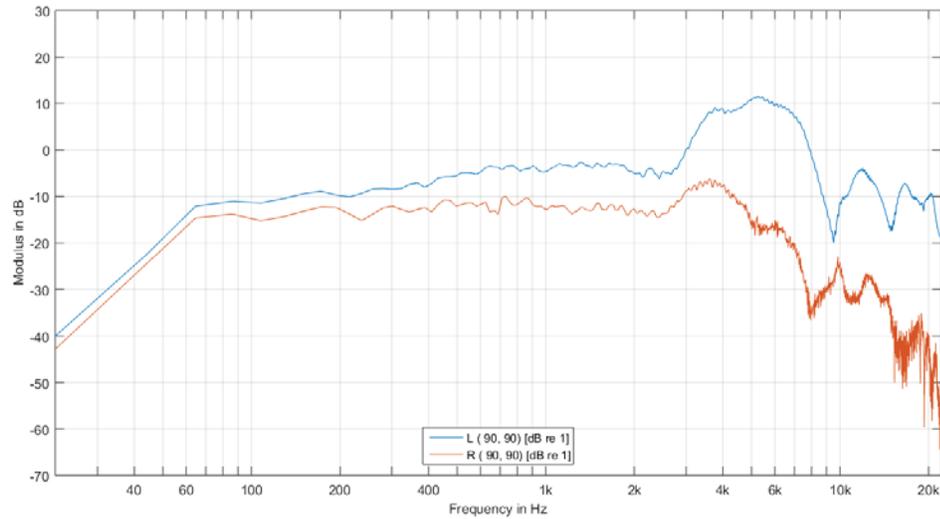


Figura A-3: Respuesta en frecuencia de "TUBerlin\_QU\_KEMAR" para una fuente ubicada al lado izquierdo del receptor.

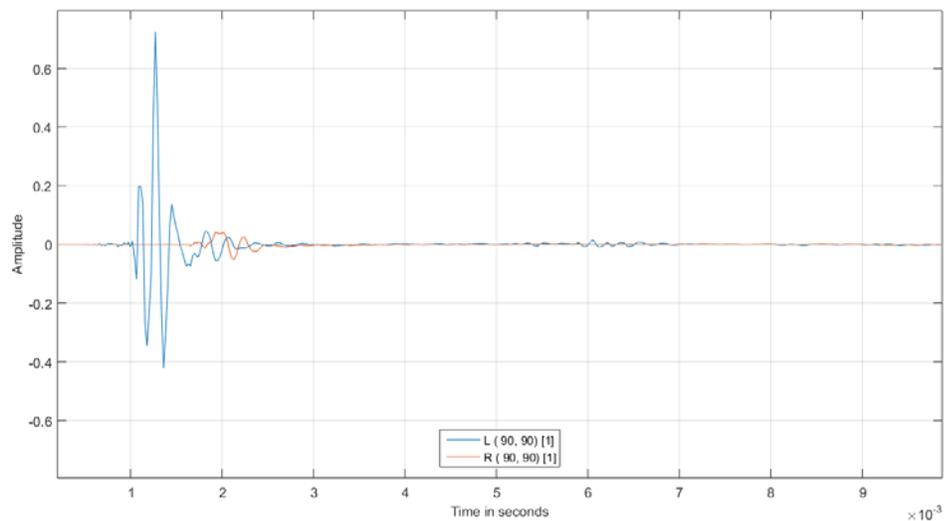


Figura A-4: Respuesta impulso de "TUBerlin\_QU\_KEMAR" para una fuente ubicada al lado izquierdo del receptor.

## A.2 Simulación de audio tridimensional de fuente sonora fija en el espacio.

```

1 %% Step 1 INICIO - VASserver
2 VASel      = 0; % 0: start VAServer.exe
3
4 deployDir = 'C:\VirtualAcoustics\VA.2018a.win32\bin';
5 useTracker = false;
6     if VASel==0
7         [~,searchresult]=system('tasklist /FI "imagename eq
8 VAServer.exe" /fo table /nh');
9         if ~strcmp(strtrim(searchresult(1:13)), 'VAServer.exe')
10            system([fullfile(deployDir, '\VAServer.exe
11 0.0.0.0:12340 '), fullfile(deployDir, 'VACoreOriginal.ini &')]);
12 pause(1) %to avoid errors
13         end
14     else
15         [~,searchresult]=system('tasklist /FI "imagename eq
16 VAGUI.exe" /fo table /nh');
17         if ~strcmp(strtrim(searchresult(1:10)), 'VAGUI.exe')
18            system([fullfile(deployDir, 'VAGUI.exe
19 '), fullfile(deployDir, 'VACore.ini
20 '), fullfile(deployDir, 'VAGUI.ini &')]);
21 pause(1) % to avoid error
22         end
23     end
24
25 %% Step 2 : Se crea un objeto y se conecta a VAserver
26 a = itaVA;
27 if ~a.get_connected % only connect if no connection to server is
28 established
29     a.connect('localhost')
30 end
31 a.reset()
32
33 a.set_output_gain(0.5); %opcional
34
35 %% Step 3: Create a receiver and assign a HRIR set
36
37 HRIRSet =
38 a.create_directivity('FABIAN_OTAH_5x5x5_256_44100Hz.v17.ir.daff'
39 ); % create a receiver and assign the HRTF set
40 L = a.create_sound_receiver('Listener');
41 a.set_sound_receiver_directivity(L, HRIRSet );
42 LHeight = 1.2; % Altura promedio de una persona sentada [m]
43 a.set_sound_receiver_position(L, [0 LHeight 0 ]) %establece la
44 posición
45 ori_initial = ita_rpy2quat(0,0,0); % calculate quaternion
46 orientation based on roll/pitch/yaw input
47 ori_initial_quat = ori_initial.e; % access quaternion
48 coefficients
49 a.set_sound_receiver_orientation(L, ori_initial_quat)
50 a.set_active_sound_receiver(L) %activar el receptor para escuchar
51 las fuentes de sonido en el estado de reproducción 'play'
52 LPosTracked = a.get_sound_receiver_position(L); % consultar la
53 posición del receptor
54 LHeightTracked = LPosTracked(2);
55
56
57
58
59
60
61
62
63
64
65
66

```

```

67 %% Step 4: Create a static virtual sound source:
68 % S1: static sound source at a defined position
69 S1 = a.create_sound_source('Source 1') % name of the sound
70 a.set_sound_source_position(S1,[0 LHeightTracked -3])
71 %va.set_sound_source_position( S, [ x y z ] )
72 Slori = ita_rpy2quat(0,0,0); % calculate quaternion orientation
73 based on roll/pitch/yaw input
74 Slori_quat = Slori.e; % access quaternion coefficients
75 a.set_sound_source_orientation(S1,Slori_quat)
76 %va.set_sound_source_orientation( S, [ a b c d ] )
77 X1 = a.create_signal_source_buffer_from_file(
78 'ruido_blanco.wav' ); % Create an audiofile signal source for
79 the sound source
80 a.set_sound_source_signal_source(S1,X1)
81 a.set_sound_source_sound_power(S1,0.05); % optionally set volume
82 of sound sourcevalue between 0 (-inf dB) and 1 (0 dB)
83 % set playback state of audiofile signal source
84 a.set_signal_source_buffer_looping( X1, true ) % looping yes/no?
85 a.set_signal_source_buffer_playback_action( X1, 'play' ) % e.g.,
86 plays the audiofile signal source
87 pause(3) % listen to the virtual scene for 3 seconds
88 a.set_signal_source_buffer_playback_action( X1, 'stop' ) %stop
89 playback
90 END
91
92
93
94
95
96
97

```

### A.3 Simulación de audio tridimensional de fuente sonora en movimiento.

```

1 %% Step 6: Create a moving virtual sound source (with
2 directivity):
3 % S2: moving virtual sound source (on a pre-defined trajectory)
4
5
6 S2 = a.create_sound_source('Source 2'); % name of the sound
7 source as string
8
9
10 % Create an audiofile signal source for the sound source (based
11 on a mono wave file)
12 X2 = a.create_signal_source_buffer_from_file( 'impulse5.wav' );
13 % ...and link the signal source to the sound source
14 a.set_sound_source_signal_source(S2,X2);
15
16
17
18 % load a directivity file in *.daff file format (e.g. directivity
19 of a trumpet)
20 S2dir = a.create_directivity( 'Singer.v17.ms.daff' );
21 % set directivity of S2
22 a.set_sound_source_directivity(S2,S2dir);
23
24
25
26 % increase sound source power due to energy loss (directivity)
27 a.set_sound_source_sound_power(S2,0.15)
28 circleR = 3; % radius of trajectory [m]
29 nlegs = 120; % number of equidistant trajectory legs
30 Tvel = 5; % time to pass nlegs points
31

```

```

32 phi_start = deg2rad(270); % start acimuth angle in [rad]
33 phi_end   = deg2rad(90);  % end acimuth angle in [rad]
34 theta     = deg2rad(0);   % zenith angle in [rad]
35
36
37 phi = linspace(phi_start,phi_end,nlegs);
38
39
40 traj_cart(:,1) = circleR*sin(phi);
41 traj_cart(:,2) = circleR*cos(phi);
42 traj_cart(:,3) = LHeight + circleR*cos(theta);
43
44
45 % set initial position of S2 (use first position of trajectory)
46 a.set_sound_source_position(S2, [traj_cart(1,1) traj_cart(1,2)
47 traj_cart(1,3)]);
48
49
50 % ... and the initial orientation
51
52 S2ori = ita_rpy2quat(0,0,0); % alternatively use ita_vu2quat
53 a.set_sound_source_orientation(S2,S2ori.e); % access quaternion
54 coefficients by .e
55
56 a.set_signal_source_buffer_looping( X2, true ); % looping yes/no?
57
58 % set period of high-precision timer [s] (for precise position
59 updates in the following update loop)
60
61 a.set_timer(Tvel/nlegs);
62
63 for idx = 1:nlegs
64     if idx==1 % start playback during first loop cycle
65         a.set_signal_source_buffer_playback_action(X2, 'play')
66     end
67     a.wait_for_timer(); % wait for a signal of the high-
68 precision timer
69
70
71     % update source position and view/up direction of S2 (virtual
72 sound source always points at receiver)
73     a.set_sound_source_position(S2, [traj_cart(idx,1),
74 traj_cart(idx,2), traj_cart(idx,3)]);
75

```

#### A.4 Simulación de audio tridimensional de dos fuentes sonoras con posición fija.

```

1 %% Step 7: Use synchronized scene actions
2
3
4 % set initial position of S2 (use first position of trajectory)
5 a.set_sound_source_position(S2, [1 0 0]);
6 a.set_sound_source_position(S1, [-1 0 0]);
7
8 % ... and the initial orientation
9 a.set_sound_source_orientation(S2,S2ori.e); % access quaternion
10 coefficients by .e
11
12
13 % shift buffer playback position of signal source
14 a.set_signal_source_buffer_playback_position(X1,0.5)
15
16
17 % everything between .lock update and .unlock update will be

```

```

18 triggered in
19 % one cycle to allow for synchronized scene events
20 a.lock_update;
21 a.set_signal_source_buffer_playback_action(X1, 'PLAY')
22 a.set_signal_source_buffer_playback_action(X2, 'PLAY')
23 a.unlock_update;
24
25
26
27 % wait until longer signal source is played back completely
28 java.util.concurrent.locks.LockSupport.parkNanos(6.5*10^9);
29
30
31 a.lock_update;
32 a.set_signal_source_buffer_playback_action(X1, 'STOP')
33 a.set_signal_source_buffer_playback_action(X2, 'STOP')
34 a.unlock_update;

```

## A.5 Simulación de audio 3D con formato SOFA.

```

1 % Start SOFA
2 SOFAstart;
3 % Carga tu respuesta de impulso en un struct
4 hrtf = SOFAload('QU_KEMAR_anechoic_3m.sofa');
5 % Muestra información sobre la respuesta al impulso
6 SOFAinfo(hrtf);
7 % Trazar una figura con la configuración de medición
8 SOFAplotGeometry(hrtf);
9
10 % Obtener información sobre la configuración de la medición
11 size(hrtf.Data.IR)
12 obtener información sobre la configuración de la medición
13 hrtf.ListenerPosition % position of dummy head
14 hrtf.SourcePosition % position of loudspeaker
15 % Orientación de cabeza ficticia + sistema de coord y unidades
16 hrtf.ListenerView
17 hrtf.ListenerView_Type
18 hrtf.ListenerView_Units
19 %Calcular la posición de origen desde el punto de vista del oyente
20 apparentSourceVector = SOFAcalculateAPV(hrtf);
21 % Escuche la HRTF con acimut de -90 °
22 apparentSourceVector(300, 1)
23 SOFAplotGeometry(hrtf, 300);
24 soundInput =
25 audioread('C:\Users\informatica\Documents\Archivos_matlab_tesis\ane
26 choic_cello.wav');
27 soundOutput = [conv(squeeze(hrtf.Data.IR(300, 1, :)), soundInput)
28 ...
29 conv(squeeze(hrtf.Data.IR(300, 2, :)), soundInput)];
30 sound(soundOutput, hrtf.Data.SamplingRate);

```

## A.6 Proceso de asignación de coordenadas a un archivo de la clase “itaHRTF”.

Para poder asignar la posición de las fuentes sonoras, con las que se realizó la medición de HRTF, dentro de los archivos de audio direccional como lo son “.daff” Y SOFA”. Se pueden utilizar las coordenadas cartesianas o esféricas. Por simplicidad se decidió utilizar las coordenadas esféricas, cuyo formato es radio, azimut y elevación ( $\rho$ ,  $\theta$ ,  $\phi$ ) (Ver figura A-5).

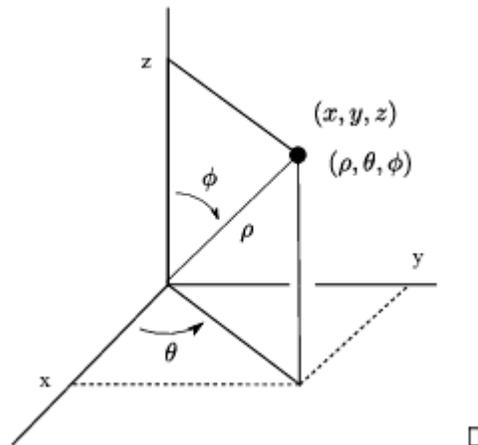


Figura A-5: Coordenadas esféricas para la descripción de la posición de la fuente de sonido.

```

1  %% Elevacion 0°
2
3
4  iC01 = itaCoordinates([ 1.00    0.00    0.00; 1.00    0.00
5  0.00], 'sph')
6  iC02 = itaCoordinates([ 1.00    0.00    0.35; 1.00    0.00
7  0.35], 'sph');
8
9
10 IR_dummy_000azimuth_000elevacion.channelCoordinates = iC01
11 IR_dummy_020azimuth_000elevacion.channelCoordinates = iC02;
12
13
14 %% Elevacion 60°
15 iC037 = itaCoordinates([ 1.00    0.52    0.00; 1.00    0.52
16 0.00], 'sph')
17 iC038 = itaCoordinates([ 1.00    0.52    0.17; 1.00    0.52
18 0.17], 'sph');
19
20
21 IR_dummy_000azimuth_060elevacion.channelCoordinates = iC037
22 IR_dummy_010azimuth_060elevacion.channelCoordinates = iC038;
23
24
25 %% Elevacion 120°
26 iC073 = itaCoordinates([ 1.00    2.09    0.00; 1.00    2.09
27 0.00], 'sph')
28 iC074 = itaCoordinates([ 1.00    2.09    0.17; 1.00    2.09
29 0.17], 'sph');
30
31
32
33

```

```
34 IR_dummy_000azimuth_120elevacion.channelCoordinates = iC073
35 IR_dummy_010azimuth_120elevacion.channelCoordinates = iC074;
36
37
38 %% Creación de archivo del tipó "itaHRTF"
39
40
41 hrtf_test2_sph =
42 itaHRTF([IR_dummy_000azimuth_000elevacion.ch(1),IR_dummy_000azimu
43 th_000elevacion.ch(2),IR_dummy_020azimuth_000elevacion.ch(1),IR_d
44 ummy_020azimuth_000elevacion.ch(2),
45 IR_dummy_000azimuth_060elevacion.ch(1),IR_dummy_000azimuth_060ele
46 vacion.ch(2),IR_dummy_020azimuth_060elevacion.ch(1),IR_dummy_020a
47 zimuth_060elevacion.ch(2),IR_dummy_000azimuth_120elevacion.ch(1),
48 IR_dummy_000azimuth_120elevacion.ch(2),IR_dummy_020azimuth_120ele
49 vacion.ch(1),IR_dummy_020azimuth_120elevacion.ch(2)])
50
```