ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

ANÁLISIS COMPARATIVO DE MÉTODOS DE INTERPOLACIÓN DE HRTFs BASADOS EN PONDERACIONES MEDIANTE SIMULACIONES EN MATLAB

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES

ANDRÉS EDUARDO ACOSTA CÓRDOVA

andres.acosta@epn.edu.ec

DIRECTOR: Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO

felipe.grijalva@epn.edu.ec

CODIRECTOR: Ph.D. ROBIN GERARDO ÁLVAREZ RUEDA

robin.alvarez@epn.edu.ec

Quito, julio 2019

AVAL

Certificamos que el presente trabajo fue desarrollado por Andrés Eduardo Acosta Córdova, bajo nuestra supervisión.

Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO
DIRECTOR DEL TRABAJO DE TITULACIÓN

Ph.D. ROBIN GERARDO ÁLVAREZ RUEDA
CODIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo, ANDRÉS EDUARDO ACOSTA CÓRDOVA, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración dejo constancia de que la Escuela Politécnica Nacional podrá hacer uso del presente trabajo según los términos estipulados en la Ley, Reglamentos y Normas vigentes.

ANDRÉS EDUARDO ACOSTA CÓRDOVA

DEDICATORIA

Dedico este trabajo primeramente a mis padres y a mi novia que siempre fueron el motor de mi vida e inspiración junto con su sacrificio para concluir dicho trabajo, a mis amigos que siempre estuvieron brindándome su apoyo y su entusiasmo, y a mis seres queridos que ya no se encuentran aquí conmigo pero que fueron parte fundamental de mi existencia.

Andrés Eduardo

AGRADECIMIENTOS

Un agradecimiento especial a mi amigo y a la vez director de tesis, Felipe Grijalva, quien siempre me proporciono su ayuda y conocimientos para poder hacer posible este trabajo y a mis amigos Gabriel, Vanessa y Jessica que siempre me ayudaron de igual forma con sus conocimientos y prestamos de equipos.

Andrés Eduardo

ÍNDICE DE CONTENIDO

AVAL I
DECLARACION DE AUTORIA II
DEDICATORIA III
AGRADECIMIENTO IV
ÍNDICE DE CONTENIDO V
RESUMEN VII
ABSTRACT VIII
1. INTRODUCCIÓN 1
1.1. Objetivo General 1
1.2. Objetivos Específicos 1
1.3. Alcance 2
1.4. Marco Teórico 2
1.4.1. Audio Espacial 3
1.4.1.1. Sistemas de Coordenadas 3
1.4.1.2. Filtros auditivos 6
1.4.1.3. Pistas para la localización del sonido 8
1.4.2. HRTF 10
1.4.2.1. Definición 10
1.4.2.2. Mediciones de las HRTFs 11
1.4.2.3. Características de las HRTFs 12
1.4.2.4. Aproximación de fase mínima de las HRTFs 15
1.4.2.5. Suavizado espectral en HRTFs 16
1.4.3. Interpolación 17
1.4.4. Métodos de interpolación HRTF [1] 17
1.4.4.1. Interpolación Lineal 18
1.4.4.2. Interpolación Bilineal 20
1.4.4.3. Interpolación Triangular Esférica 22
1.4.4.4. Interpolación Splines Cúbicos 22
1.4.5. Distorsión Espectral 23

2. METODOLOGÍA 24
2.1. Creación de la base de datos general 24
2.1.1. Obtención de los archivos SOFA 24
2.1.2. Exportación y compilación de archivos SOFA al Matlab 24
2.1.3. Creación de la base de datos general 26
2.1.4. Preprocesamiento de las HRIRs 27
2.2. Algoritmos de interpolación para HRTFs 27
2.2.1. Algoritmo para la interpolación lineal adyacente 29
2.2.1.1. Algoritmo de interpolación lineal por azimut 29
2.2.1.2. Algoritmo de interpolación lineal por elevación 30
2.2.2. Algoritmo para la interpolación bilineal 33
2.2.3. Algoritmo para la interpolación triangular esférica 37
2.2.4. Algoritmo para la interpolación por splines cúbicos 40
3. RESULTADOS Y DISCUSIÓN 41
3.1. Análisis del efecto del suavizado espectral 41
3.2. Análisis del efecto de los pasos de azimut y elevación 44
3.3. Análisis de la distorsión espectral en los sujetos 45
3.4. Análisis de la distorsión espectral en función de la dirección 47
4. CONCLUSIONES Y RECOMENDACIONES 54
4.1. CONCLUSIONES 54
4.2. RECOMENDACIONES 55
5. REFERENCIAS BIBLIOGRÁFICAS 56
ANEXOS 59

RESUMEN

El audio espacial o audio en 3D, son técnicas que tiene como principal objetivo la simulación de sonido, provenientes de fuentes sonoras, las cuales se encuentran ubicadas en una determinada posición en el espacio. Esto es posible gracias a las Funciones de Transferencia Derivadas de la Cabeza (*Head-Related Transfer Functions*,HRTF), las cuales modelan características anatómicas del sujeto y su interacción con el campo sonoro incidente.

En este trabajo, se realiza un análisis comparativo de cuatro métodos de interpolación para HRTFs mediante la métrica de la distorsión espectral y la utilización de parámetros tales como número de HRTFs interpolantes, suavizado espectral, sujetos y direcciones de las fuentes sonoras. Por otra parte, este proyecto también propone una nueva metodología basada en PCA (*Principal Component Analysis*) y razones de áreas para los métodos de interpolación bilineal y triangular esférico de modo que se adecuen a las bases de datos de HRTFs medidas en mallas no uniformes.

Finalmente, los principales resultados basados en la métrica de la distorsión espectral, muestran que dicha métrica se reduce considerablemente al utilizar el parámetro del suavizado espectral. Por otro lado, la metodología propuesta basada en PCA para interpolaciones bilineal y triangular esférica es una alternativa viable de en mallas no uniformes ya que su exactitud comparada con la interpolación spline es similar en términos de distorsión espectral.

PALABRAS CLAVE: Audio espacial, HRTF, métodos de interpolación, distorsión espectral, suavizado espectral.

ABSTRACT

Spatial audio or 3D audio are techniques whose main objective is to simulate sound sources located in arbitrary positions in space. This is possible thanks to the Head-Related Transfer Functions (HRTFs), which model the anatomical characteristics of the subject and their interaction with the incident sound field.

In this work, we compare four interpolation methods for HRTFs using spectral distortion as a metric for parameters such as number of interpolating HRTFs, spectral smoothing, subjects and addresses of the sound sources. On the other hand, this project also proposes a new methodology based on Principal Component Analysis (PCA) and area ratios for the bilinear and triangular spherical interpolation methods so that they might be applied to HRTFs databases measured in non-uniform meshes.

Finally, our results show that the spectral distortion is considerably reduced after applying spectral smoothing. On the other hand, the proposed methodology based on PCA for bilinear and triangular spherical interpolations is a viable alternative in non-uniform mesh since its accuracy when compared to spline interpolation is similar in terms of spectral distortion.

KEYWORDS: spatial audio, HRTF, smoothing, interpolation methods,spectral distorsions

1. INTRODUCCIÓN

El audio espacial binaural tiene como objetivo simular fuentes de sonido en una ubicación arbitraria específica mediante la utilización de las Funciones de Transferencia Derivada de la Cabeza (*Head-Related Transfer Function*, HRTF) [1]. La HRTF describe el filtrado espectral de una fuente de sonido causado por la cabeza, el pabellón auditivo y el torso, antes de alcanzar el tímpano [2].

Las HRTFs son funciones de valores complejos que contienen varias características, las cuales definen nuestra forma de localizar fuentes sonoras, tales como la diferencia de tiempo interaural (*Interaural Time Difference*, ITD), la diferencia de nivel interaural (*Interaural Level Difference*, ILD) y características espectrales. Estas señales estáticas, en conjunto con las señales dinámicas (e.g. movimientos de la cabeza), definen nuestra percepción de audio en 3 dimensiones [3].

Las mencionadas funciones de transferencia suelen estar disponibles en bases de datos, las cuales fueron obtenidas a partir de mediciones en personas en una cámara anecóica siendo estos valores muestreados a cierta resolución espacial (e.g. cada 5 grados). Cabe recalcar que este procedimiento de medición es lento, costoso y complejo por los equipamientos requeridos [4].

El problema surge cuando se desea obtener la función de transferencia de una dirección que no existe en la base de datos para lo cual se utilizan métodos de interpolación espacial. Ante esta problemática el presente proyecto tiene como propósito fundamental realizar un análisis comparativo entre los cuatro métodos de interpolación de HRTFs basadas en ponderaciones.

1.1. Objetivo General

El objetivo general de este estudio técnico es realizar un análisis comparativo de métodos de interpolación de HRTFs basados en ponderaciones mediante simulaciones en Matlab.

1.2. Objetivos Específicos

- Estudiar los conceptos básicos de audio espacial y HRTFs.
- Analizar los métodos de interpolación de HRTFs basados en ponderaciones.
- Implementar los métodos de interpolación de HRTFs basados en ponderaciones.
- Comparar los cuatro métodos de interpolación implementados mediante la utilización de la métrica de distorsión espectral.

1.3. Alcance

Los métodos de interpolación más sencillos estiman una HRTF en una posición desconocida como un promedio ponderado de las HRTF medidas cercanas. Se han utilizado diferentes esquemas para elegir los pesos de dicha ponderación, como interpolación lineal, interpolación bilineal, splines cúbicos, entre otras variantes [1].
Este proyecto realiza una comparación de los siguientes métodos de interpolación espacial de HRTFs [1]:

- Interpolación lineal adyacente (adjacent linear interpolation): Consiste en obtener el azimut objetivo de una determinada HRTF de una zona azimutal conformada por azimuts vecinos.
- Interpolación bilineal (Bilinear interpolation): en esta interpolación se utiliza una cuadrícula de mediciones cuyos vértices son las direcciones más cercanas a la dirección objetivo y obtiene dicha dirección a través de sumas ponderadas.
- Interpolación triangular esférica (Spherical triangular interpolation): Determina la HRTF de la dirección objetivo a partir de las 3 posiciones más cercanas a dicha dirección.
- Interpolación spline cúbica (Cubic spline interpolation): Consiste en utilizar polinomios cúbicos segmentados para que se adapten a las HRTFs.

Para comparar los cuatro métodos mencionados, interpolaremos HRTFs de bases de datos en formato SOFA [2]. Específicamente, se varian los siguientes parámetros para realizar el análisis comparativo:

- Suavizado espectral para HRTFs
- Tamaño del conjunto de HRTFs interpolantes (i.e. HRTFs de posiciones conocidas) utilizado para la interpolación de HRTFs en posición desconocidas.

Por otro lado, se utilizan métricas de distorsión espectral para comparar la HRTF interpolada con la HRTF real. Cabe recalcar que el presente trabajo interpolará HRTFs de fase mínima por lo que la comparación será realizada en magnitud. Finalmente, los algoritmos serán implementados en Matlab.

1.4. Marco Teórico

En esta sección se explica las bases y conceptos fundamentales de audio espacial e interpolación.
En la Sección 1.4.1 el cual explica la teoría entorno al audio en el espacio y los tipos de sistemas de coordenadas que existen. Luego la Sección 1.4.2 presenta los conceptos, características y las maneras de medición para las funciones de transferencia relacionadas con la cabeza (HRTF). A continuación la Sección 1.4.3 hace referencia sobre la definición de interpolación y como se la realiza. Por último, en la Sección 1.4.4 y 1.4.5 se describe el

proceso de interpolación para cada uno de los cuatro métodos de interpolación para HRTFs implementadas en este proyecto y a su vez el concepto de la distorsión espectral.

1.4.1. Audio Espacial

El audio espacial o también denominado audio tridimensional (3D) consiste en modelar las propiedades antropométricas del ser humano (i.e. oído, cabeza, torso) en campos sonoros incidentes.
Existen diferentes maneras de generar el audio espacial, tales como *Panning [5], Binaural, Sound Field Synthesis [6] y HRTF [7]*. Para el análisis del presente estudio técnico se utiliza el audio espacial generado a través de HRTFs.

1.4.1.1. Sistemas de Coordenadas

Nuevamente, se menciona que en el audio espacial, toda fuente de sonido debe tener una dirección y una distancia con respecto a la cabeza del quien lo escuche. A su vez es esencial decir que el origen de los sistemas de coordenadas en audio espacial es por lo general en el centro de la cabeza a nivel de los oídos.
Estos sistemas de coordenadas utilizan 3 planos de referencia, los cuales proporcionan una mejor orientación al momento de ubicar una fuente de sonido. Estos planos son [1]:

- Plano horizontal: Este plano está conformado por dos vectores que apuntan hacia la dirección frontal y hacia la derecha o izquierda.
- Plano mediano: Plano que apunta hacia sus direcciones frontal y superior.
- Plano latera o vertical: El plano está conformado por un vector que apunta hacia la dirección superior y por otro vector que apunta hacia izquierda o derecha.

Todos estos planos son perpendiculares entre sí [1], y para obtener una mejor comprensión, en la Figura 1.1 se aprecia detalladamente dichos planos.

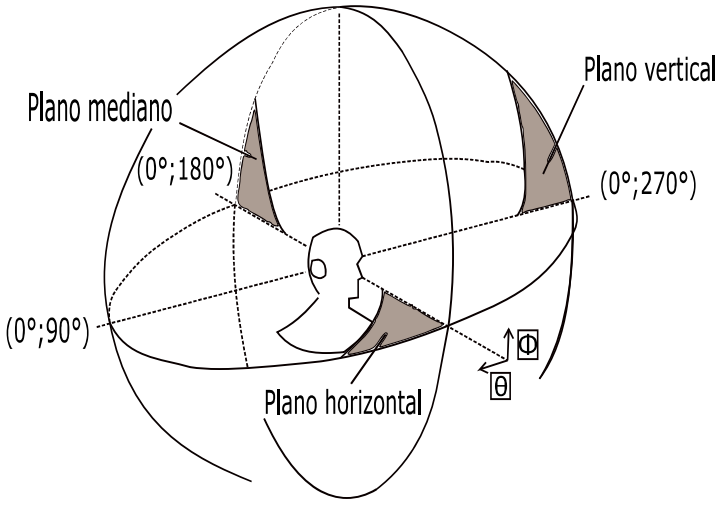

Figura 1.1: Planos de referencia [8]

A continuación, se describen 2 sistemas de coordenadas ampliamente utilizados en audio espacial: sistema de coordenadas esféricas y sistema de coordenadas interaurales [1]. A menos que se indique lo contrario, en este trabajo se utiliza como sistema de coordenadas predeterminado al sistema esférico.

- En el **sistema de coordenadas esféricas** mostrado en la Figura 1.2, para ubicar una determinada fuente sonora se utilizan las siguientes coordenadas (r,θ, ϕ), donde r es la distancia desde la fuente de sonido hasta el centro de la cabeza que es el origen; ϕ es la elevación que es un ángulo formado entre un vector que va desde el origen del sistema de coordenadas hacia la fuente sonora y el plano horizontal cuyo rango de valores es $-90° \leq \phi \leq 90°$; θ es el azimut, el cual es un ángulo medido entre la proyección horizontal del vector y el eje Y (en este caso la dirección frontal), cuyos valores van de $0° \leq \theta \leq 360°$.

  Con respecto a las referencias del presente sistema de coordenadas en el plano horizontal se tiene que [1]:

  - $\theta = 0°$ es la dirección frontal.
  - $\theta = 90°$ es la dirección izquierda.
  - $\theta = 180°$ es la dirección posterior.
  - $\theta = 270°$ es la dirección derecha.

  Estas referencias del sistema de coordenadas esféricas se resumen fácilmente en la Figura 1.3.

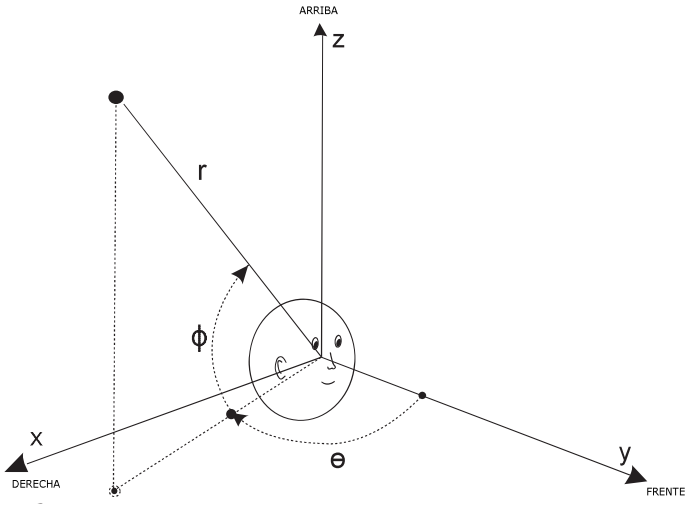

Figura 1.2: Sistema de coordenadas esféricas [1].

- De igual manera el **sistema de coordenadas polar interaurales** mostrado en la Figura 1.4, localiza las fuentes sonoras con las coordenadas (r, Θ, Φ). Sin embargo, Θ y Φ no son los mismos ángulos de azimut y elevación que en el sistema de coordenadas anteriormente mencionado, ya que Θ es el ángulo de azimut polar interaural, el cual

está formado entre el vector direccional y el plano medio, mientras que Φ es el ángulo de elevación polar interaural, medido entre la proyección del vector direccional al plano mediano y su eje y (dirección frontal). En cuanto al rango de estos ángulos con respecto al anterior sistema, son diferentes ($-90° \leq \phi \leq 270°$ y $-90° \leq \theta \leq 90°$). Por otro lado las referencias para el sistema de coordenadas polar interaural son las siguientes [1]:

- (Θ, Φ)=$(0°, 0°)$ es la dirección frontal.
- (Θ, Φ)=$(0°, 90°)$ es la dirección superior.
- (Θ, Φ)=$(0°, 180°)$ es la dirección posterior.
- (Θ, Φ)=$(0°, 270°)$ es la dirección inferior.
- (Θ, Φ)=$(90°, 0°)$ son las direcciones derecha e izquierda.

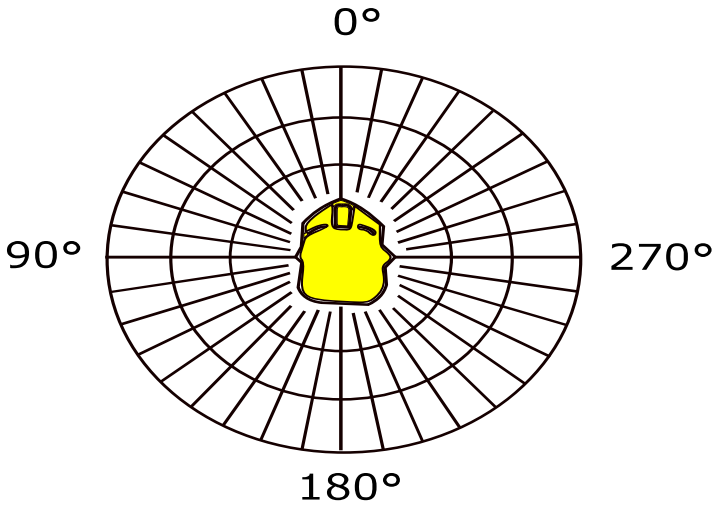

Figura 1.3: Referencia para Sist. Coord. Esféricas.

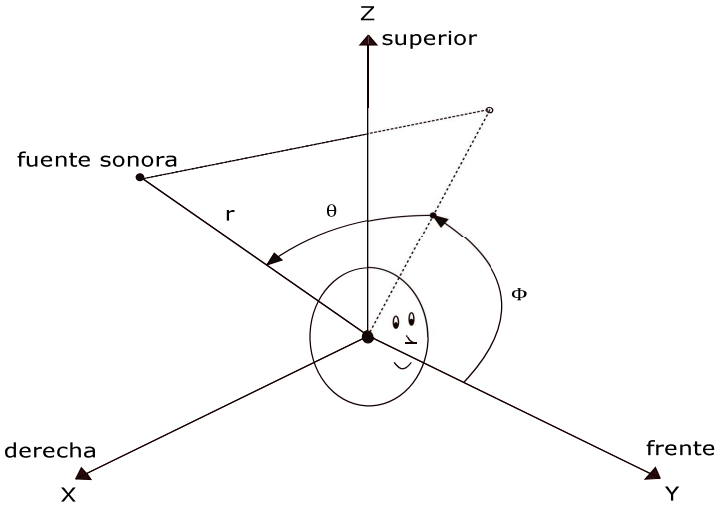

Figura 1.4: Sistema de coordenadas polar interaural [1].

1.4.1.2. Filtros auditivos

El oído esta conformado por tres partes que son: la parte externa conformada por el pabellón auditivo y su conducto auditivo; la parte media conformada por el tímpano y huesecillos (martillo, yunque y estribo) y por último el oído interno conformado por la ventana oval, ventana circular y la cóclea (caracol) (Figura 1.5).

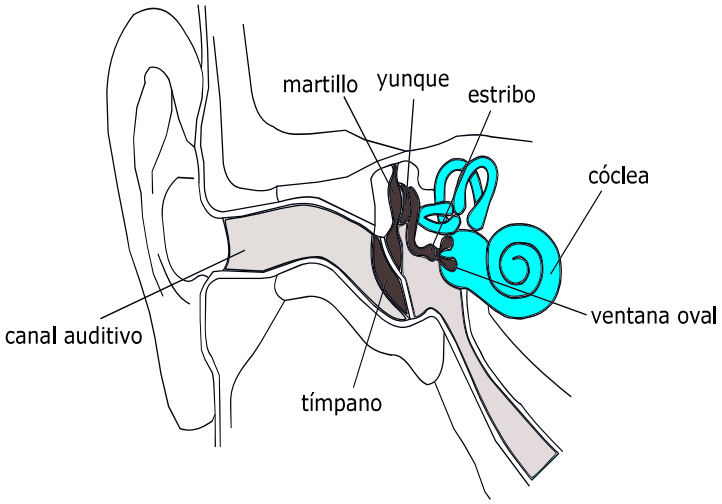

Figura 1.5: Estructura del oído [9]

El sistema auditivo funciona de la siguiente manera: el sonido entra por el canal auditivo hasta llegar a una membrana llamada tímpano. De manera que se produzcan vibraciones en la membrana, las cuales causan movimientos en los huesecillos del oído medio que transmiten información de frecuencia y amplitud. Por consiguiente, son enviadas a la ventana oval mediante el estribo que actúa como pistón para que estas vibraciones lleguen al caracol.
En cuanto al caracol o cóclea es una parte del sistema auditivo humano que se halla en el oído interno, el cual actúa como un analizador de frecuencia [1]. Esto es posible gracias a la membrana basilar que se encuentra en el interior de la cóclea, cuya longitud es idéntica a la longitud de la cóclea desenrollada y su funcionamiento se puede resumir de la siguiente manera: al momento que llegan las vibraciones al caracol, entran a la membrana basilar y estas recorren hasta una determinada ubicación de la misma, donde al llegar a dicha ubicación le corresponderá una frecuencia característica, en pocas palabras, cada ubicación de la membrana tiene un diferente valor de frecuencia, en donde las frecuencias más bajas se ubican en el ápice del caracol (final de la membrana basilar) y las más altas en la base del mismo (inicio de la membrana basilar) [1]. Todo lo mencionado se sintetiza en la Figura 1.6.
Un problema que tiene la membrana basilar es cuando detecta dos sonidos simultáneamente, donde la diferencia entre sus frecuencias es pequeña y en consecuencia, el patrón de respuesta de la membrana se vuelve compleja. Por tanto, una solución para el presente inconveniente, es el enmascaramiento de un tono por un ruido pasa banda. Este enmascaramiento tiene como objetivo hacer inaudible al sonido de nuestro interés. Para conseguirlo, se debe aplicar ruido con un determinado ancho de banda, de manera que oculte la fre-

cuencia del sonido a hacer inaudible. Como resultado se tiene un nuevo ancho de banda resultante conocido como ancho de banda crítico en la frecuencia central (*critical bandwidth at the center frequency*) [10].
Este fenómeno es producto de la característica principal de la membrana basilar que es la de ser un analizador de frecuencia ya que, como dijimos antes, la membrana responderá a una frecuencia característica dependiendo de su ubicación. En consecuencia, la membrana se puede considerar finalmente como un banco de filtros pasa banda o también llamado banco de filtros auditivos [1].

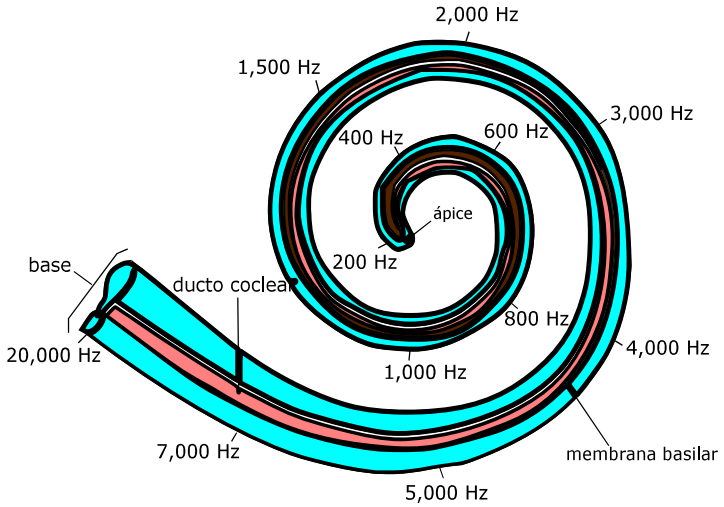

Figura 1.6: Análisis de frecuencia en membrana basilar [11].

En investigaciones pasadas, Fletcher modeló los filtros auditivos como filtros rectangulares [10]. Puesto que la resolución en frecuencia del sistema auditivo está relacionada con la forma y el ancho de los filtros auditivos, si el ancho de banda del ruido de enmascaramiento está dentro del ancho de banda efectivo del filtro auditivo, el ruido enmascara efectivamente el tono a la frecuencia característica. Por otro lado, si el ancho de banda del ruido de enmascaramiento es más ancho que el ancho de banda efectivo, los componentes del ruido fuera del ancho de banda efectivo del filtro auditivo tienen poco efecto sobre el enmascaramiento. Además, el ancho de banda crítico proporciona una aproximación del ancho de banda de los filtros auditivos [1].
Por otro lado, algunos estudios [12] han determinado que ciertos filtros auditivos reales no tienden a ser simétricos. Sin embargo, se los puede aproximar con filtros pasa banda de banda rectangulares simétricos. Estos filtros tienen una transmisión en su pasa banda igual a la transmisión máxima del filtro auditivo especificado, y transmite la misma potencia de ruido blanco que el filtro auditivo original [1]. Por último al ancho de banda de este filtro se lo llama ancho de banda rectangular equivalente (ERB) [12].
La fórmula para calcular el ERB es la siguiente:

$$ERB = 24{,}7(4{,}37 * f + 1), \tag{1.1}$$

donde f, es la frecuencia central o característica en Khz.

1.4.1.3. Pistas para la localización del sonido

El objetivo fundamental de la localización del sonido es conocer la ubicación de la fuente sonora por intermedio de dos parámetros que son la distancia y su dirección con respecto al oyente. Por tanto, al relacionar la ubicación de un sonido con nuestra percepción direccional del sonido, esta se encontrará integrada por 3 pistas que están clasificadas como: pistas binaurales, dinámicas y espectrales.

Pistas binaurales

Las pistas de localización son fundamentales al momento de localizar una determinada fuente sonora en el plano horizontal que incluyen las diferencias relativas entre las ondas sonoras que llegan a los oídos. Estas diferencias fueron descritas por la teoría duplex de Rayleigh [13] como la diferencia de tiempo interaural (*Interaural Time Difference*,ITD) y la diferencia de nivel interaural (*Interaural Level Difference*,ILD).

La ITD es la diferencia en la medición del tiempo en que se demoran las ondas sonoras en llegar,desde una fuente de sonido hasta los oídos de la persona. Por otra parte, esta diferencia de tiempo depende tanto de la frecuencia como de la dirección de donde proviene el sonido. [14].

La ITD tiene mayor peso en frecuencias por debajo de 700 Hz, ya que en esta banda las dimensiones de la cabeza humana son más pequeñas que la longitud de onda sonora de la fuente de sonido (frecuencia es inversamente proporcional a la longitud de onda sonora). De esta manera, el sistema auditivo detecta las diferencias de retardo de fase sin confusión [1]. Por otro lado, a altas frecuencias (es decir, desde 1.5 kHz), el ITD se vuelve ambiguo porque la percepción de la posición lateral ya no es proporcional a la diferencia de fase percibida. Aun así, la ITD a altas frecuencias se considera una señal secundaria, debido a que el sistema auditivo puede extraer las diferencias de retardo interaural de las envolventes de las ondas de sonido [15].

Por otro lado, la diferencia de nivel interaural, se da cuando la fuente de sonido se desplaza del plano mediano. En consecuencia, la presión sonora que emite dicha fuente se atenúa en el oído más lejano a la misma, mientras que en el oído más cercano sucede lo contrario puesto que en este escenario la cabeza actúa como un obstáculo. En conclusión, a esta diferencia de presiones sonoras es a lo que se denomina diferencia de nivel interaural (ILD). La ILD tiene mayor importancia en frecuencias superiores a 1.5KHz. Por lo cual al presentarse frecuencias altas, las longitudes de onda del sonido tendrán un tamaño menor a las dimensiones de la cabeza humana promedio, i.e., en las frecuencias donde la ITD tiende a tener ambigüedades [16].

Pistas dinámicas

En cuanto a las pistas de localización para fuentes sonoras, se ha hablado de 2 (ITD e ILD), las cuales no resultan ser totalmente eficientes al momento de localizar el origen del

sonido ya que en la realidad es posible que exista un conjunto de ILDs e ITDs idénticos para diferentes localizaciones de fuentes sonoras. Por ejemplo, al considerar un sistemas de coordenadas interaurales donde se mantiene el azimut interaural constante. En este escenario se tendrá una superficie cónica cuyos valores de ITD e ILD serán teóricamente idénticos, tal como se observa en la Figura 1.7.

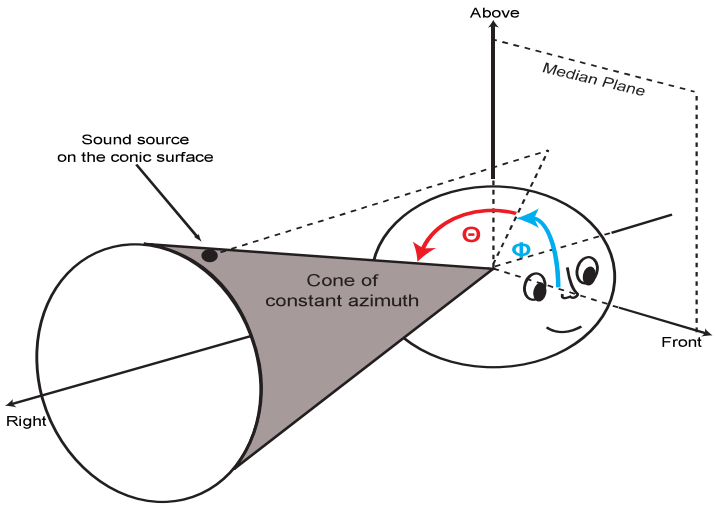

Figura 1.7: Cono de confusión [1].

A esta superficie cónica se lo denomina "*Cono de confusión*". Una manera de reducir las ambigüedades producidas por esta superficie cónica es a través de las pistas dinámicas como por ejemplo los movimientos de cabeza [17, 18], ya que al realizar un pequeño movimiento ya sea adelante, atrás, izquierda o derecha, se puede percibir fácilmente la localización de la fuente sonora. Para una mejor comprensión de lo mencionado se lo ilustrará mediante la Figura 1.8.

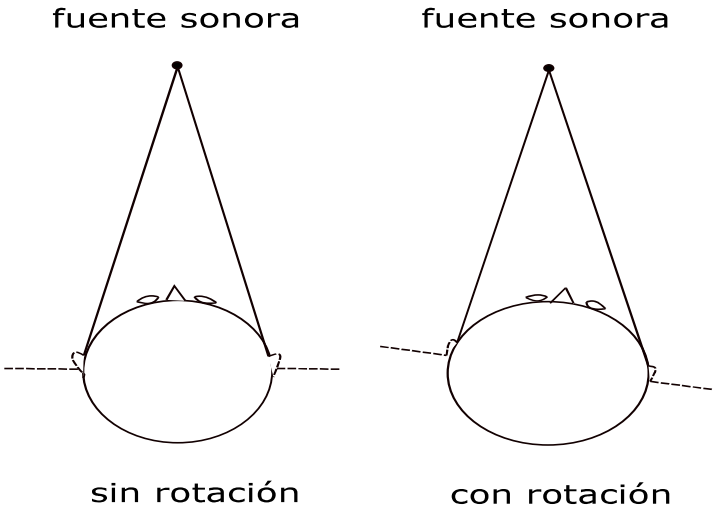

Figura 1.8: Movimientos de cabeza [1].

Pistas espectrales

Estas pistas son conocidas como pistas monoaurales y son otro tipo de pistas de localización auditiva, causada por características anatómicas tales como el oído externo, el cual esta conformado por el pabellón auricular (oreja), cabeza y torso. Por lo que se refiere a la geometría de la oreja (pabellón auditivo), es una característica muy importante ya que el tamaño del pabellón auditivo varia en cada persona y esto nos interesará especialmente en frecuencias superiores a 3 kHz cuando su tamaño es comparable a la longitud de onda de la fuente [19]. Por tal motivo, dichas diferencias espectrales son utilizadas por el sistema auditivo para resolver el cono de confusión. Además estas pistas son codificadas en las funciones de transferencia relacionadas con la cabeza (HRTF) que se mencionan en la siguiente sección. Por último es importante destacar que las pistas espectrales son las más importantes al momento de localizar las elevaciones de una determinada fuente sonora [20].

1.4.2. HRTF

Las funciones de transferencia relacionadas con la cabeza (*Head Related Transfer Functions*, HRTF) son piezas importantes al momento de entrar en el análisis del audio espacial binaural, i.e., son aquellas que nos permiten ubicar un sonido en una determinada posición deseada en el espacio. En esta sección se enfoca en el concepto de las HRTFs, así como su contraparte en tiempo, i.e., las respuestas de impulso relacionadas con la cabeza (*Head Related Impulse Response*, HRIRs).

1.4.2.1. Definición

El sonido emitido por una fuente sonora en el campo libre llega a ambos oídos después de interactuar con las características anatómicas del sujeto (cabeza, torso y pabellón auditivo). Por consiguiente, la señal resultante compuesta por varias pistas descritas anteriormente tales como la ITD, la ILD y las pistas espectrales, que juntas se modelan a través de las HRTF. Así, se tiene 2 HRTFs, uno para el oído izquierdo y otro para el derecho, i.e., H_L y H_R respectivamente, las cuales se definen como [1]

$$\begin{aligned} H_L(r,\theta,\phi,f,a) &= \frac{P_L(r,\theta,\phi,f,a)}{P_0(r,f)}, \\ H_R(r,\theta,\phi,f,a) &= \frac{P_R(r,\theta,\phi,f,a)}{P_0(r,f)}, \end{aligned} \tag{1.2}$$

donde,

P_L, P_R : Son las presiones sonoras del oído izquierdo y derecho respectivamente en función de la frecuencia.

P_o : Presión sonora en función de la frecuencia de campo libre en el centro de la cabeza con la ausencia de ella [1].

a : Variable que depende únicamente de las propiedades anatómicas del sujeto, generalmente son mediciones antropométricas de la cabeza humana.

r : Distancia entre la fuente sonora y la cabeza a nivel de los oídos.
Así mismo las HRTFs pueden como no depender de la distancia (r) y para ello tienen dos nombres. En el caso de que la HRTF no dependa de la distancia se llama *HRTF de campo lejano*. Por otro lado las que si dependen se denominan *HRTF cercanas*. El presente estudio técnico se enfoca exclusivamente a las HRTFs de campo lejano.

1.4.2.2. Mediciones de las HRTFs

Para comprender mejor el proceso de medición, en la Figura 1.9 muestra como se debe preparar el lugar en donde se realizan las mediciones de HRTFs. Se observa que lo primero es estar en una cámara semi-anecóica ya que es muy importante evitar las reflexiones de los sonidos. Después hay que situar al sujeto (oyente) en el centro de la cámara y alrededor de él, colocar amplificadores en diferentes direcciones (ubicaciones) para luego reproducir sonidos. Dicho lo anterior, se procede a mencionar los pasos para llevar a cabo la obten-

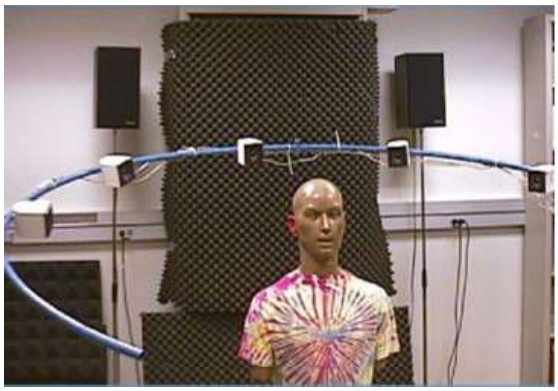

Figura 1.9: Colocación de micrófonos, amplificadores y el oyente (KEMAR) [21].

ción de HRTFs que se resumen en la Figura 1.10. La obtención se lo realiza mediante la reproducción de una señal analítica (es decir, una señal sintética conocida) en la dirección deseada (a una distancia de al menos 1 m de la fuente para las HRTF de campo lejano). Luego se procede a la medición de las respuestas de impulso relacionadas con la cabeza (*HRIR*) sin ningún procesamiento a través de micrófonos de sonda ubicados en los oídos (en el canal auditivo). Conviene subrayar que todo el procedimiento es para una determinada dirección por lo que se debe repetir los pasos para cada dirección que se requiera.

Para el diagrama de bloques de la Figura 1.10, es importante previamente considerar los siguientes aspectos para su eficiente funcionamiento:
Señales analíticas de entrada: Es muy común que en sistemas de medición de audio exista siempre la presencia del ruido. Es por eso que las propiedades de la señal deben tener una relación señal a ruido (SNR) altas. Por lo tanto, lo más simple es aumentar la potencia a través de dispositivos como amplificadores, etc. Sin embargo el gran problema es que se producen distorsiones, de manera que se viola la linealidad de los sistemas electroacústicos. Por tanto, para las HRIRs una posible solución es utilizar señales que contengan un factor de cresta bajo (razón entre valores pico y efectivo) [22]. Así por ejemplo, las señales de

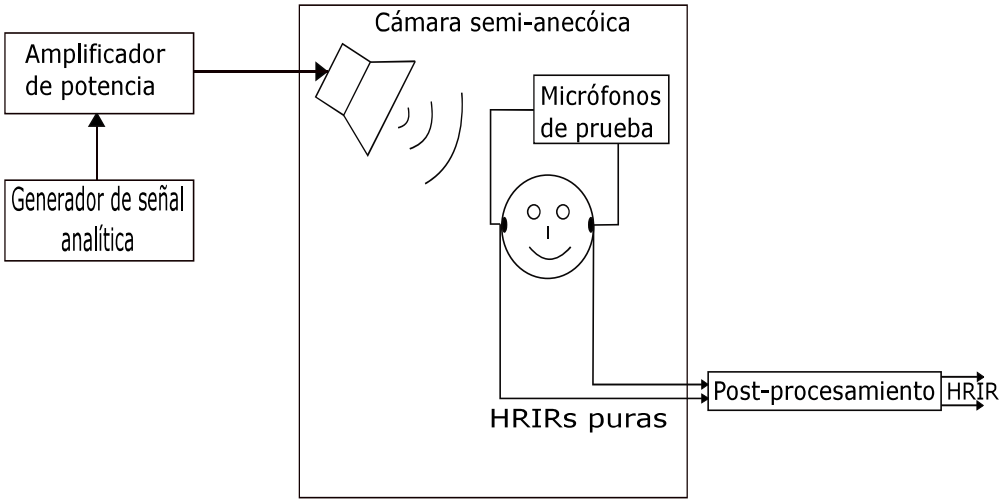

Figura 1.10: Diagrama de bloques para medir HRIRs.

barridos sinusoidales [23] han demostrado ser las mejores para la generación de respuestas de impulso ya que generan las HRIRs con alto SNR y manteniendo a la vez un bajo factor de cresta [24].

Ubicación de los micrófonos: La ubicación más empleada para la colocación de estos micrófonos son en las entradas a los oídos internos (entrada del conducto auditivo) por mejor seguridad y por tener las mediciones lo más reales posibles. Es decir, que los micrófonos sean capaces de simular una audición común y corriente del oyente.

Cámara semi-anecóica: Generalmente se debe utilizar una cámara anecóica ya que tiene la ventaja de eliminar las reflexiones de los sonidos que se produzcan para las mediciones. Sin embargo el gran problema de estas cámaras son sus dificultades técnicas y elevados costos. Por lo cual las mediciones de las respuestas de impulso se las hace normalmente en una cámara semi-anecóica [25]. Aunque estas no eliminan las reflexiones completamente, se puede controlar el tiempo en que las ondas sonoras reflejadas lleguen a los oídos a través de materiales absorbentes acústicos del cual esta hecha la cámara [1].

Por otra parte, todas las mediciones son almacenadas en bases de datos. Por lo que, para el presente estudio técnico se utilizan bases de datos SOFA, las cuales son formatos espacialmente orientado para acústica y como lo dice su nombre, sirve para almacenar bases de datos acústicos espaciales [2], tales como las mismas HRTFs y HRIRs. Cabe mencionar, como información adicional que las bases de datos que posee el formato SOFA solo contiene las respuestas de impulso HRIR. Además, dichas bases de datos contienen las localizaciones (direcciones) en coordenadas esféricas para cada una de las fuentes sonoras.

1.4.2.3. Características de las HRTFs

Características en tiempo

Al hablar específicamente en el dominio del tiempo, la señal más importante en el tiempo son las ITD. Aunque hay varios métodos para estimar la ITD (para más detalles leer [14]), un

enfoque directo es calcular la diferencia entre los retardos de inicio entre el HRIR izquierdo y derecho.
En conclusión y en términos matemáticos, se define a la diferencia de tiempo interaural como

$$ITD(\theta, \phi) = t_L - t_R \tag{1.3}$$

donde t_L, t_R = Son los retardos, es decir, los tiempo en que se demoran en llegar el sonido a sus respectivos oídos.
El siguiente ejemplo (Figura 1.11) explica el calculo de la ITD, como se mencionó anteriormente.

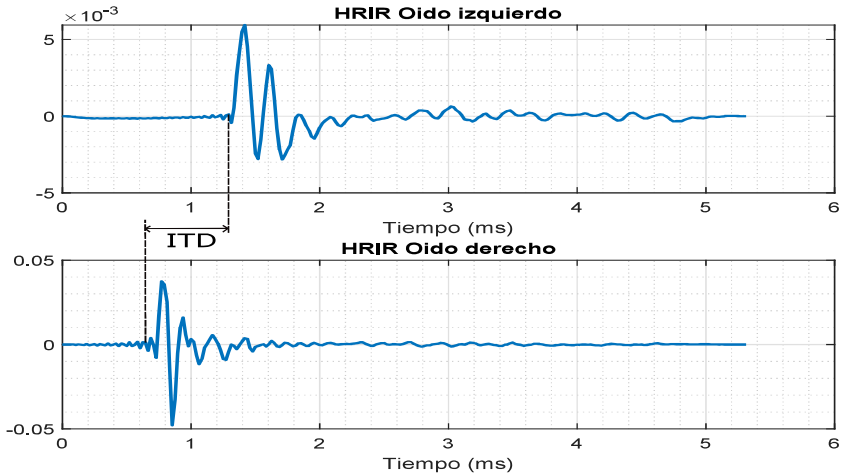

Figura 1.11: Ejemplo de calculo ITD con $\theta = 90°$ y $\phi = 0°$.

Características en frecuencia

A continuación para describir las propiedades de las HRTFs con relación a la frecuencia, se los muestra a través de la Figura 1.12 que son representaciones gráficas de las magnitudes HRTFs medidas en 4 ángulos de azimuts diferentes en el plano horizontal. Por lo que se refiere al gráfico, se tiene las siguientes observaciones:

- En frecuencias menores a los 200Hz, existen atenuaciones en el nivel de la señal. Ya que los amplificadores tienen limitaciones en sus respuestas a bajas frecuencias.
- Entre los 400 y 500Hz sucede que la atenuación se aproxima a 0, debido a que las HRTFs son independientes de la frecuencias.
- Como se explico anteriormente a frecuencias mayores a 1.5KHz tiene mayor peso las ILD antes que las ITD. Por tanto las ILD son visibles, i.e., la diferencia entre magnitudes HRTFs entre los oídos son apreciables a frecuencias superiores a los 1.5KHz.

También a la diferencia de nivel interaural, matemáticamente, se la define como:

$$ILD(\theta, \phi) = 20log_{10}\frac{|H_R(\theta, \phi, f)|}{|H_L(\theta, \phi, f)|} \tag{1.4}$$

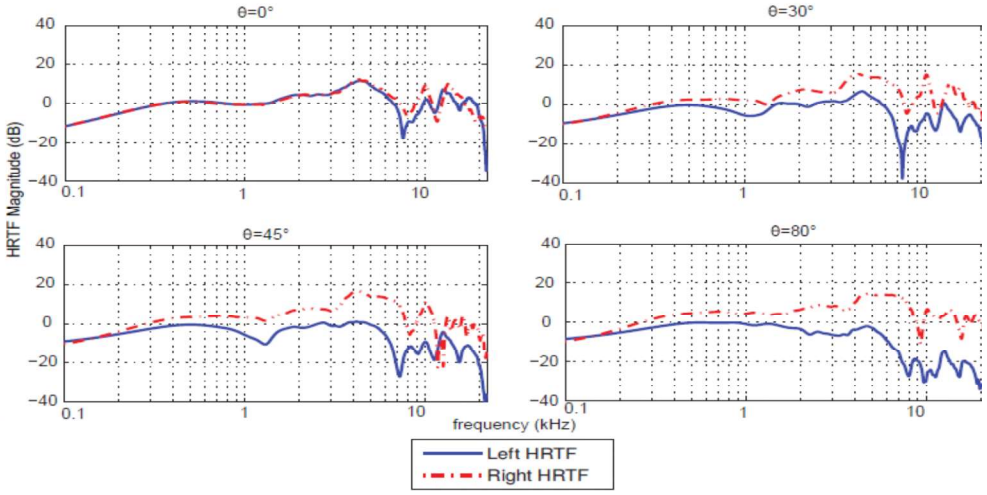

Figura 1.12: Magnitudes HRTF vs frecuencia en determinados azimuts.

Por último, con relación a las magnitudes de HRTF, estas son diferentes en cada persona ya que se debe recordar que en cada sujeto las propiedades anatómicas no son las mismas que en otros individuos. Esto porque el sonido al llegar al oído externo, su señal será filtrada a través de sus pabellones auditivos, lo que a su vez causa respuestas de impulsos diferentes, y por consiguiente las HRTFs también lo son.

Para ilustrar mejor la explicación anterior, la Figura 1.13 indica las diferencias en las HRTFs para cada individuo.

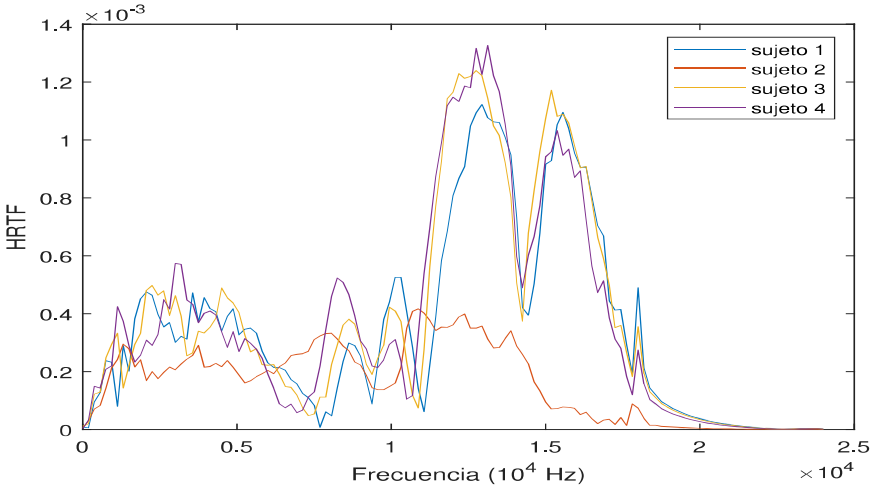

Figura 1.13: Magnitudes HRTFs en diferentes sujetos.

Bases de datos de HRTFs

Utilizando procedimientos para mediciones de HRTFs como los mencionados en la Sección 1.4.2.2 u otros procesos semejantes [23], varios grupos de investigación han creado bases de datos de HRTFs las cuales se encuentran disponibles de manera pública. Por ejemplo,

una de las bases de datos más completas es la base de datos CIPIC [25] porque tiene mediciones para un gran número de individuos (43 sujetos y KEMAR) además de sus datos antropométricos. Se realizaron 1250 mediciones en coordenadas interaurales para 25 azimuts no uniformemente distribuidos con una resolución máxima de $5°$ para ubicaciones cercanas al plano medio y 50 elevaciones con resolución de $5{,}625°$ en el rango de $-45°$ a $230{,}625°$. Una desventaja de la base de datos CIPIC es su baja resolución en direcciones laterales (i.e., $15°$ a $20°$). Otra base de datos popular es la base de datos ARI [26]. Las principales ventajas de ARI sobre la base de datos CIPIC es que la primera ofrece HRTF de más sujetos (aproximadamente 150 sujetos) y tiene una mejor resolución espacial. También hay bases de datos que solo midieron HRTF para KEMAR [27] o maniquíes similares. Para una lista no exhaustiva de bases de datos HRTF, se recomienda la lectura de la referencia [3]. Por otro lado, en un esfuerzo por facilitar el intercambio de datos entre investigadores y usuarios, el *Audio Engineering Society (AES)* ha estandarizado el Formato Orientado Espacialmente para la Acústica [28]. SOFA es un formato de archivo para almacenar e intercambiar datos acústicos orientados espacialmente, como las HRTF. Actualmente, las bases de datos HRTF más populares están disponibles públicamente para descarga en formato SOFA [2], incluidas las bases de datos CIPIC y ARI.

1.4.2.4. Aproximación de fase mínima de las HRTFs

Las HRTFs provienen de las HRIRs, esto debido a que las HRIRs dependen del tiempo mientras que las HRTFs de la frecuencia. En consideración a lo anterior, para el análisis del presente estudio se lo realiza en función de la frecuencia, es por eso que se aplica la Transformada de Fourier, de manera que se obtenga las HRTFs.

Como resultado de dicha transformada las funciones de transferencia tienen valores complejos [1]. Por consiguiente, la HRTF estará compuesta por una magnitud y una fase. De manera generalizada, las funciones de transferencia tales como las HRTFs se las puede representar como una descomposición [29] entre el producto de su función de fase mínima $H_{min}(\theta, \phi, f)$, una función del tipo pasa-todo (i.e. *all-pass function*) $exp[j\Psi_{all}(\theta, \phi, f)]$ de ganancia unitaria y una función de fase lineal $exp[-2j\pi fT(\theta, \phi)]$, de tal forma que :

$$H(\theta, \phi, f) = H_{min}(\theta, \phi, f) \cdot exp[j\Psi_{all}(\theta, \phi, f)] \cdot exp[-2j\pi fT(\theta, \phi)] \tag{1.5}$$

donde,

$T(\theta, \phi)$ es el tiempo de retraso producido por la propagación de las ondas sonoras antes de que lleguen a los oídos y

Ψ es la fase de la señal.

Por otro lado, la magnitud de una función de transferencia y su correspondiente magnitud de fase mínima son iguales [29], es decir, $|H| = |H_{min}|$, por consiguiente se tiene que:

$$H_{min}(\theta, \phi, f) = |H(\theta, \phi, f)| \cdot exp[j\Psi_{min}(\theta, \phi, f)] \tag{1.6}$$

donde la fase de la función de fase mínima y el logaritmo de la magnitud están relacionados por la transformada de Hilbert

$$\Psi_{min}(\theta, \phi, f) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\ln |H(\theta, \phi, x)|}{f - x} \cdot dx \tag{1.7}$$

Con respecto a la Ecuación 1.5 y en base a lo explicado anteriormente se deduce que la fase de las HRTFs esta dada por:

$$\Psi(\theta, \phi, f) = \Psi_{min}(\theta, \phi, f) + \Psi_{all}(\theta, \phi, f) - 2\pi f T(\theta, \phi) \tag{1.8}$$

Ahora, si se toma en cuenta que la componente de fase de la función pasa-todo (all-pass) es despreciable, la HRTF puede aproximarse como:

$$H(\theta, \phi, f) \approx |H(\theta, \phi, f)| \cdot exp[-j2\pi f T(\theta, \phi, f)] \tag{1.9}$$

La ecuación 1.9 se denomina la aproximación de fase mínima de una HRTF, en donde la HRTF se aproxima por su función de fase mínima en cascada con un retardo puro o una función de fase lineal $T(\theta; \phi)$ [1]. En la práctica, este retraso es solo la ITD calculada a partir del método descrito en la Sección 1.4.2.3 u otro [14]. Vale la pena mencionar que, en este caso, el retraso se colocaría solo en la HRTF (izquierda o derecha) que garantiza que el tiempo de inicio sea siempre positivo. Los estudios realizados por [30] y [31] han demostrado la validez del modelo de fase mínima de HRTF. Esta aproximación es importante porque permite procesar la HRTF utilizando solo su magnitud. Una vez que se procesa la magnitud de la HRTF, la HRTF compleja se puede reconstruir utilizando la Ecuación 1.9. Sin embargo, para el presente proyecto, solo se usa la magnitud de las HRTFs para aplicar los métodos de interpolación, los cuales se explican en la Sección 1.4.4.

1.4.2.5. Suavizado espectral en HRTFs

En cuanto a lo explicado en la Sección 1.4.2.3, las HRIR de una persona se trasladan a su dominio en frecuencia (HRTFs) y estas a su vez son diferentes en cada sujeto. Un aspecto importante en la medición de las funciones de transferencia es el oído, en razón de que tienen diferentes formas y dimensiones, tales que permiten percibir sonidos a determinadas frecuencias.

Un punto crítico de estas mediciones de HRIRs es cuando se filtran con ruido. Por lo cual al momento de trasladar al dominio de la frecuencia, se presentan un sin número de picos tales que complican el proceso de interpolación. Es así que se utiliza el suavizado espectral (smoothing), el cual funciona como un filtro auditivo que permite eliminar la mayor cantidad posible de picos producto del ruido en las HRTFs (Figura 1.14).

Para implementar el suavizado espectral en las HRTFs, se utiliza el concepto del ERB, en razón de que usa un ancho de banda rectangular en una determinada banda para enmas-

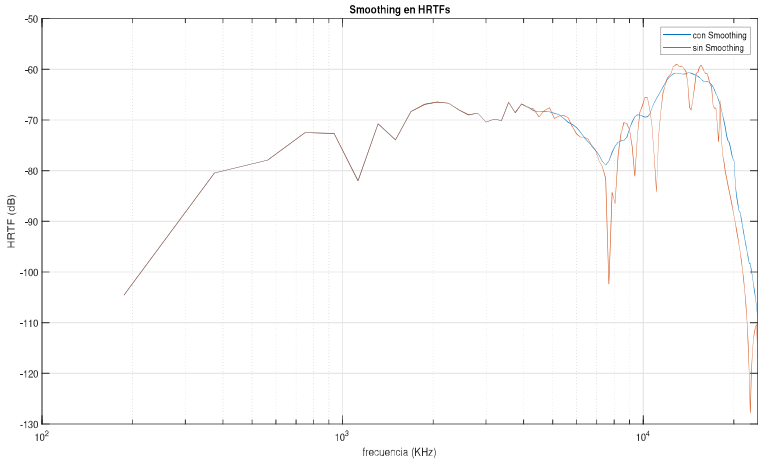

Figura 1.14: HRTFs con y sin suavizado espectral.

carar los picos de frecuencia producidos posiblemente por el ruido. A fin de que se tenga una curva prolongada o suavizada como se vio en la Figura 1.14. Finalmente, para evaluar la validez del suavizado en HRTFS se realizó experimentos psicoacústicos [32] para determinar la audibilidad del detalle espectral de las HRTFs en alta frecuencia en donde se demuestra que [1] las HRTFs pueden ser suavizadas con anchos de banda de 1 ERB (equivalent rectangular bandwidth) y 2 ERB para bajas (f <5 kHz) y altas (f ≥ 5 kHz) frecuencias, respectivamente, sin que el oyente pierda percepción auditiva.

1.4.3. Interpolación

La interpolación es un método matemático (o método numérico) que se utiliza para determinar un valor desconocido entre 2 o más valores ya conocidos, o a su vez es conocido como un modelo matemático que sirve para realizar pronósticos [33]. La Figura 1.15 es un ejemplo simple del concepto de interpolación, en donde se observa la diferencia entre una función que pasa por 3 puntos (azul) que es la original y la función interpolada (negro).
Un parámetro muy importante a considerar en el proceso de la interpolación es la distancia en que se encuentra el punto objetivo (i.e. el punto que se desea interpolar) con respecto a sus puntos vecinos más cercanos (i.e. los puntos conocidos) . Es decir, que mientras más cerca se encuentre el objetivo de los vecinos, mejor es la interpolación (i.e. mejor aproximación) caso contrario, la interpolación del punto objetivo esta lejos del valor exacto.

1.4.4. Métodos de interpolación HRTF [1]

En esta Sección se explica a detalle cada uno de los cuatro métodos de interpolación para HRTFs que forman parte del alcance de este proyecto: lineal adyacente (por azimut y elevación), bilineal, triangular esférica y splines cúbicos. Es importante mencionar que en el contexto de audio espacial los puntos objetivo y vecinos se refieren a la magnitud de las HRTFs.

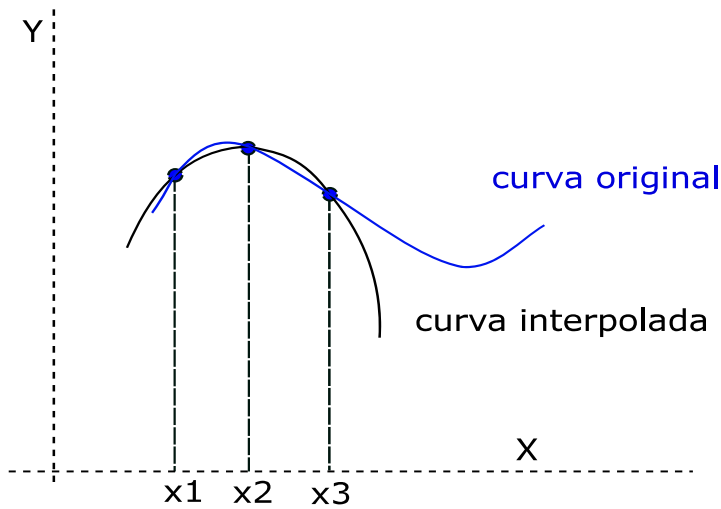

Figura 1.15: Ejemplo interpolación (original vs interpolada) [34].

1.4.4.1. Interpolación Lineal

En lo referente a este método, este tiene dos variantes, es decir, que se puede hacer la interpolación por elevación ó por azimut.

Interpolación lineal por azimut : Esta interpolación consiste en interpolar una HRTF desconocida cuya ubicación se encuentra entre dos puntos vecinos ubicados en una misma elevación, tal como se ve en la Figura 1.16. Para realizar su interpolación esta se basa en multiplicar las HRTFs de los puntos vecinos por unos valores denominados pesos, que son razones entre distancias azimutales. A continuación, las formulas de los pesos se detallan en la Ecuación 1.10 y Ecuación 1.11.

$$A_{i+1} = \frac{\theta - \theta_i}{\theta_{i+1} - \theta_i} \tag{1.10}$$

$$A_i = 1 - \frac{\theta - \theta_i}{\theta_{i+1} - \theta_i} \tag{1.11}$$

Para el uso de las Ecuaciones 1.10 y 1.11 se debe considerar que $\theta_i < \theta < \theta_{i+1}$.
donde
A_i, A_{i+1} : Son los pesos de los dos vecinos.
θ_i, θ_{i+1} : Son los ángulos de azimut de los dos vecinos.
θ : Es el ángulo del azimut objetivo.

Finalmente estas multiplicaciones son sumadas (Ecuación 1.12) para determinar la HRTF en ese punto objetivo (i.e. $\hat{H}$, la HRTF interpolada o aproximada).

$$\hat{H}(\theta, f) \approx A_{i+1} H(\theta_{i+1}, f) + A_i H(\theta_i, f) \tag{1.12}$$

donde
$\hat{H}$: Es la HRTF interpolada en el punto objetivo (azimut objetivo).
$H(\theta_{i+1}, f), H(\theta_i, f)$: Son las HRTFs de los azimuts vecinos.

Para ilustrar mejor el método de la interpolación lineal adyacente por azimut, la Figura 1.17 muestra como se debe seleccionar los puntos vecinos y objetivos, previo al cálculo de los pesos para su debida interpolación.

Interpolación lineal por elevación: De igual forma esta interpolación consiste en interpolar una HRTF desconocida cuya ubicación se encuentra entre dos puntos vecinos que están en el mismo azimut, tal como se ve en la Figura 1.18. Igualmente en esta interpolación se calcula los pesos como en la lineal adyacente por azimut, solo que en vez de usar el azimut, esta vez se usa la elevación (Ecuaciones 1.13 y 1.14).

$$A_{i+1} = \frac{\phi - \phi_i}{\phi_{i+1} - \phi_i} \tag{1.13}$$

$$A_i = 1 - \frac{\phi - \phi_i}{\phi_{i+1} - \phi_i} \tag{1.14}$$

Donde
$\phi_i < \phi < \phi_{i+1}$
A_i, A_{i+1} : Son los pesos de los ángulos de elevación de los dos vecinos.
ϕ_i, ϕ_{i+1} : Son los ángulos de elevación de los dos vecinos.
ϕ : Es la elevación objetiva.

Una vez calculado sus pesos, se procede a multiplicarlos por sus respectivas HRTFs vecinas, de manera que la HRTF interpolada se calcule aproximadamente tal como se muestra en la Ecuación 1.15.

$$H\hat{R}TF(\phi, f) \approx A_{i+1}H(\phi_{i+1}, f) + A_iH(\phi_i, f) \tag{1.15}$$

Donde:
$H\hat{R}TF$: Es la HRTF interpolada en el punto objetivo.
$H(\phi_{i+1}, f), H(\phi_i, f)$: Son las HRTFs de las elevaciones vecinas.

De igual forma, la figura 1.19 es un ejemplo que nos indica la manera de como seleccionar las HRTFs vecinas y objetivas cuando se interpole linealmente por elevación.

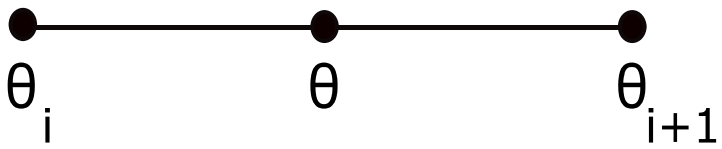

Figura 1.16: Interpolación lineal adyacente por azimut.

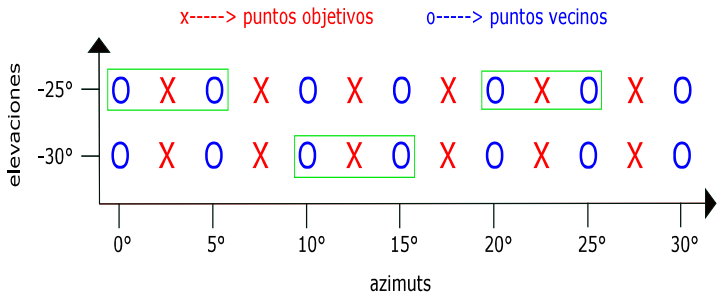

Figura 1.17: Ejemplo interpolación lineal azimut.

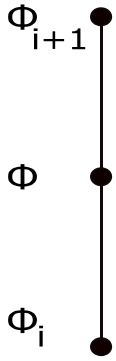

Figura 1.18: Interpolación lineal adyacente por elevación.

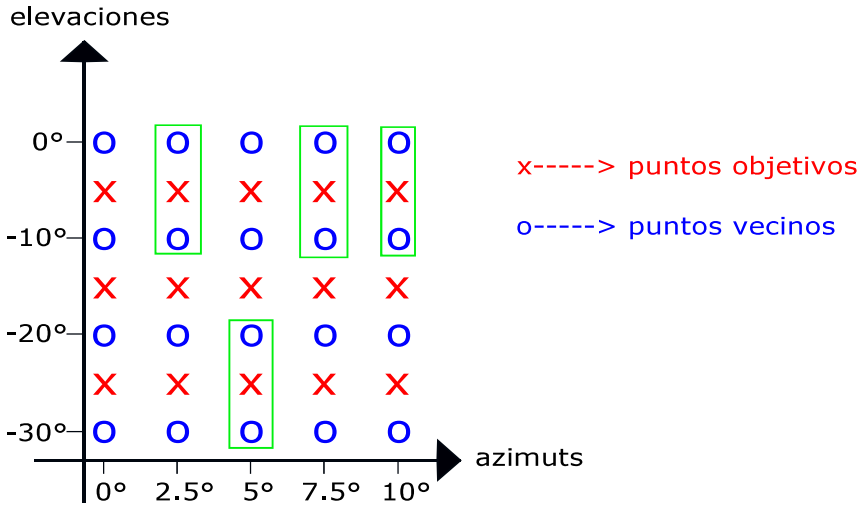

Figura 1.19: Ejemplo interpolación lineal por elevación.

1.4.4.2. Interpolación Bilineal

En este tipo de interpolaciones, su manera de encontrar las HRTFs desconocidas es diferente a la lineal adyacente ya que no se utiliza solo los vecinos que estén en la misma elevación o azimut, i.e., utiliza ambos para encontrar sus cuatro puntos vecinos más cercanos. Idealmente los vecinos forman un cuadrado en el caso de que la malla en la que las HRTFs fueron medidas sea uniforme (ver Figura 1.20).

Donde

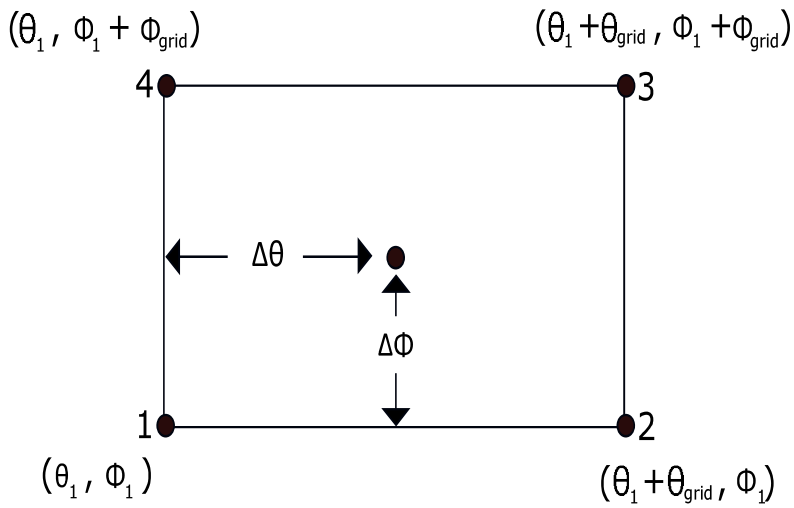

Figura 1.20: Ilustración de la interpolación bilineal [1].

θ_{grid} : es la distancia entre dos puntos vecinos adyacentes en una misma elevación.
ϕ_{grid} : es la distancia entre dos puntos vecinos adyacentes en un mismo ángulo de azimut.
$\Delta\theta$: es la distancia desde el punto objetivo (punto dentro del cuadrado) hasta el lado vertical izquierdo del cuadrado.
$\Delta\phi$: es la distancia desde el punto objetivo (punto dentro del cuadrado) hasta el lado horizontal inferior del cuadrado.
Para esta interpolación, se hace uso de sumas ponderadas, es decir, primero se calcula los pesos, cuyas formulas se muestran en la Ecuación 1.16 para cada uno de los cuatro puntos vecinos.

$$\begin{aligned} A_1 &= (1-A_\theta)(1-A_\phi) & A_2 &= A_\theta(1-A_\phi) \\ A_3 &= A_\theta A_\phi & A_4 &= (1-A_\theta)A_\phi \\ A_\theta &= \frac{\Delta\theta}{\theta_{grid}} & A_\phi &= \frac{\Delta\phi}{\phi_{grid}} \end{aligned} \tag{1.16}$$

Donde
A_1, A_2, A_3, A_4 son los pesos de cada uno de los puntos vecinos, respectivamente.
Luego, estos pesos son multiplicados por las HRTFs de esos mismos puntos para finalmente sumarlas y obtener la HRTF interpolada en el punto objetivo (punto dentro del cuadrado).
La fórmula para calcular la HRTF interpolada o aproximada, es la que se muestra en la Ecuación 1.17.

$$\hat{H}(\theta,\phi,f) \approx A_1H(1,f) + A_2H(2,f) + A_3H(3,f) + A_4H(4,f) \tag{1.17}$$

donde:
$H(1,f), H(2,f), H(3,f), H(4,f)$ son las HRTFs de cada uno de los puntos vecinos, respectivamente.

1.4.4.3. Interpolación Triangular Esférica

Esta interpolación es muy similar a la bilineal, tiene la misma característica de buscar a sus cuatro vecinos más cercanos. Dicho esto, una vez encontrado los vecinos, procede a determinar cual de los puntos es el más cercano con respecto al punto objetivo para luego escoger los dos puntos adyacentes con respecto al punto más cercano. Es decir, solo utiliza tres de los cuatro puntos para su interpolación. Considerando que la malla en la que se midió las HRTFs sea unirforme, estos tres puntos vecinos forman un triangulo que encierra al punto objetivo como se muestra en la Figura 1.21, donde $\theta_{grid}, \phi_{grid}, \Delta\theta, \Delta\phi$, son las mismas definiciones que en el método de interpolación bilineal.

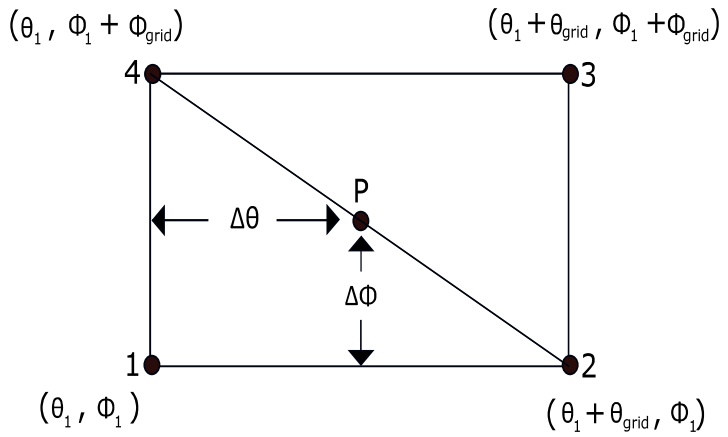

Figura 1.21: Ilustración de la interpolación triangular esférica [1].

De igual forma esta interpolación también utiliza ponderaciones (pesos) que se calculan con la Ecuación 1.18 las cuales se observan a continuación.

$$\begin{aligned} A_1 &= (1 - A_\theta - A_\phi) \\ A_2 &= A_\theta \\ A_4 &= A_\phi \\ A_\theta &= \frac{\Delta\theta}{\theta_{grid}} = \frac{|\theta - \theta_1|}{\theta_{grid}} \\ A_\phi &= \frac{\Delta\phi}{\phi_{grid}} = \frac{|\phi - \phi_1|}{\phi_{grid}} \end{aligned} \tag{1.18}$$

Después, estos pesos son multiplicados con las HRTFs de los tres puntos vecinos, respectivamente para finalmente ser sumados y obtener las HRTFs desconocidas (HRTF aproximada) en el determinado punto objetivo mediante la Ecuación 1.19.

$$\hat{H}(\theta, \phi, f) \approx A_1 H(1, f) + A_2 H(2, f) + A_4 H(4, f) \tag{1.19}$$

1.4.4.4. Interpolación Splines Cúbicos

La interpolación por splines cúbicos es uno de los más sencillos de implementar ya que la herramienta de programación MATLAB tiene una función destinada para hacer uso de este

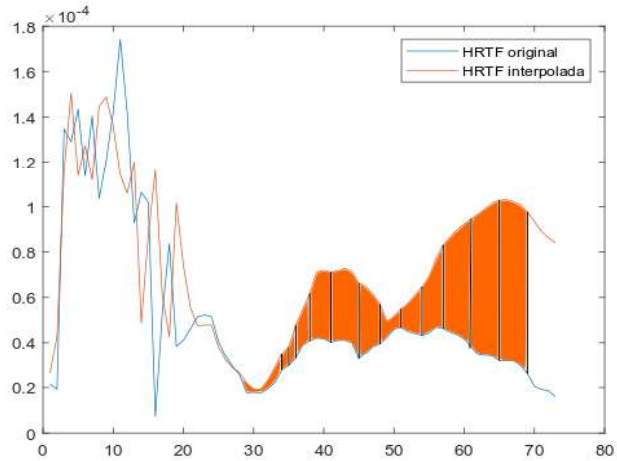

Figura 1.22: Ejemplo de la distorsión espectral (zona naranja).

modelo matemático. Con respecto a esta interpolación, consiste en determinar un polinomio de tercer orden que sea capaz de pasar por dos puntos vecinos cercanos al punto objetivo. De manera que una vez se determine ese polinomio, únicamente solo se ingrese las ubicaciones de donde queremos conocer las HRTFs, junto con las HRTFs y ubicaciones de los puntos vecinos y este nos devuelva las HRTFs interpoladas. En este caso las ubicaciones son los azimuts y las elevaciones que son nuestras direcciones del sistema de coordenadas esféricas.

1.4.5. Distorsión Espectral

Al concluir el proceso de interpolación de las HRTFs es necesario saber que tan aproximado resulta las HRTFs interpoladas con respectos a las originales. Por tanto se utiliza la distorsión espectral o también comúnmente llamada *spectral distortion*, el cual es una métrica que permite calcular el margen de error que existe entre las dos funciones de transferencia.
A continuación, se presenta una gráfica (Figura 1.22) que muestra la diferencia entre una HRTF real y la interpolada, i.e., se refiere a la zona de color naranja.
En cuanto a la manera de calcularla, primeramente las unidades al ser del tipo adimensional en la distorsión espectral se las expresa en dB. La fórmula que utiliza se basa en la raíz cuadrada del error cuadrático medio que se calcula con la ecuación 1.20.

$$SD_M = \sqrt{\frac{1}{N_f}\sum_{k=1}^{N_f}\left(20log_{10}\frac{|H(k)|}{|\hat{H}(k)|}\right)^2} \tag{1.20}$$

donde:
N_f, es el numero de componentes de frecuencia.
$H(k)$, son las HRTFs originales.
$\hat{H}(k)$, son las HRTFs interpoladas.

2. METODOLOGÍA

El presente capítulo tiene como objetivo explicar de que manera o a través de que procedimientos fueron posibles la implementación de los métodos de interpolación en HRTFs.
La Sección 2.1 se refiere únicamente a la creación de la base de datos principal que es utilizada para aplicar los métodos de interpolación. Por otro lado, la Sección 2.2 explica como fueron creados los 4 métodos de interpolación en Matlab.

2.1. Creación de la base de datos general

Para esta Sección se detalla todos los procesos necesarios para la creación de una base de datos que contiene información de todas las direcciones y HRTFs de todos los sujetos. La Sección 2.1.1 describe los pasos para obtener los archivos en formato SOFA, los cuales sirvieron en la construcción de la base de datos general. Mientras que la Sección 2.1.2 explica la exportación y compilación de dichos archivos a través del Matlab. Luego la siguiente Sección (2.1.3) trata sobre como unificar toda la información extraida de cada uno de los archivos del formato SOFA. Por último la Sección 2.1.4 indica las restricciones que se deben aplicar a las HRTFs de todos los sujetos previo a la interpolación.

2.1.1. Obtención de los archivos SOFA

En esta sección se explica los pasos para obtener los archivos del formato SOFA [2], los cuales contienen información esencial para cada sujeto y estas son: sus direcciones (azimut y elevación) en coordenadas esféricas y las HRIRs.
Con respecto a dichos archivos, se los obtiene a través del siguiente enlace web: http://sofacoustics.org/data/database/ari/. Luego, al ingresar a dicha página web, se encuentran los archivos en formato SOFA (Figura 2.1) los cuales se deben descargar para posteriormente exportarlos al Matlab. Además hay que resaltar que cada uno de los archivos que se encuentran en el enlace representa a una base de datos completa para un solo sujeto. En otras palabras, existen 89 archivos o bases de datos que corresponden a 89 sujetos respectivamente.

2.1.2. Exportación y compilación de archivos SOFA al Matlab

Para empezar con la exportación de los archivos SOFA, es necesario tener todos los archivos ya descargados en el computador para luego almacenar toda la información en una sola carpeta que la llamaremos dtf_nhs.
Luego de que los archivos ya fueron guardados en el programa, se utiliza un Script llamado inicio.m el cual nos permite compilar cada uno de los archivos SOFA para poder extraer su contenido de manera individual. El código del Script es el siguiente:

Segmento de código 2.1: Script inicio.m

```
SOFAstart; %carga todas las librerias necesarias del formato SOFA
```

Index of /data/database/ari

Name	Last modified	Size	Description
Parent Directory		-	
dtf_nh2.sofa	2014-05-23 11:39	2.9M	
dtf_nh4.sofa	2014-05-23 11:39	2.9M	
dtf_nh5.sofa	2014-05-23 11:39	2.9M	
dtf_nh8.sofa	2014-05-23 11:40	2.9M	
dtf_nh10.sofa	2014-05-23 11:40	2.9M	
dtf_nh11.sofa	2014-05-23 11:40	2.9M	
dtf_nh12.sofa	2014-05-23 11:40	2.9M	
dtf_nh13.sofa	2014-05-23 11:40	2.9M	
dtf_nh14.sofa	2014-05-23 11:40	2.9M	
dtf_nh15.sofa	2014-05-23 11:40	2.9M	
dtf_nh16.sofa	2014-05-23 11:40	2.9M	

Figura 2.1: Bases de datos formato SOFA [2].

```
%cargo a la variable hrtf todo el contenido de un determinado sujeto (archivo)
hrtf=SOFAload('dtf_nh2.sofa'); %se carga el archivo cuyo nombre es ...
    dtf_nh2.sofa
```

Realizado todos los pasos anteriores, se procede a la compilación del código para que posteriormente devuelva una estructura de datos de un determinado archivo (sujeto) que se almacena en la variable hrtf, la cual está creada y ubicada en el workspace del Matlab.
Al abrir la variable hrtf, el cual es una estructura de datos, se encuentran 47 campos de entre los cuales solo nos interesan 2; el campo SourcePosition y el campo Data (Figura 2.2). Si se desea una descripción completa de los 47 campos, esta se encuentra detallada en el ANEXO B.

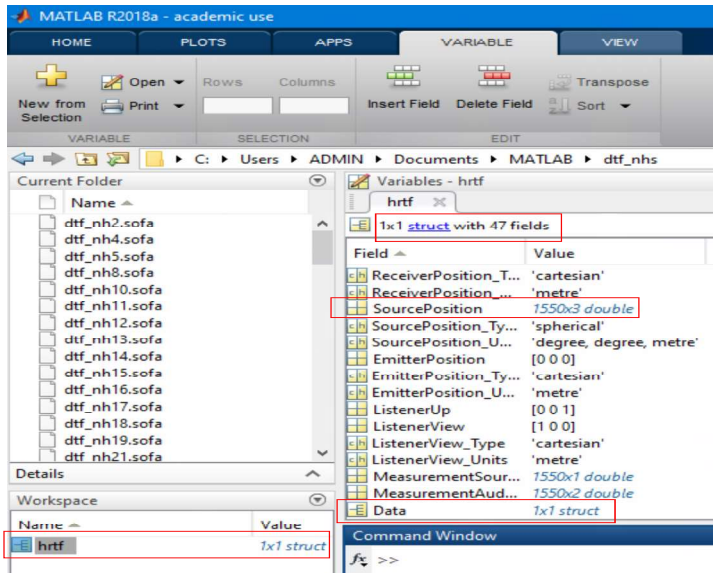

Figura 2.2: Contenido variable hrtf.

El campo SourcePosition como se muestra en la Figura 2.2 es una matriz cuyas dimensiones

Tabla 2.1: Contenido del campo Data.

Campo	Valor	Descripción
IR	[1550x2x256]	Contiene las HRIRs medidas.
SamplingRate	48000	Frecuencia de muestreo.
SamplingRate_Units	'Hertz'	Unidades de muestreo de la frecuencia de muestreo.
Delay	[0 0]	Tiempo de llegada en muestras para cada receptor (*Time of Arrival*, TOA).

son 1550 filas por 3 columnas, la cual contiene todas las direcciones de las fuentes sonoras para un solo sujeto (Figura 2.3). Además estas direcciones son las mismas para todos los sujetos y se encuentran en coordenadas esféricas.

Variables - hrtf.SourcePosition

hrtf | hrtf.SourcePosition

hrtf.SourcePosition

	1	2	3
1	0	-30	1.2000
2	180	-25	1.2000
3	0	-20	1.2000
4	180	-15	1.2000
5	0	-10	1.2000
6	180	-5	1.2000
7	0	0	1.2000
8	180	5	1.2000
9	0	10	1.2000
10	180	15	1.2000
11	0	20	1.2000
12	180	25	1.2000

Figura 2.3: Contenido campo SourcePosition.

Por otra parte, el campo Data es otra estructura de datos cuyo contenido se muestra en la Tabla 2.1. De manera que para esta estructura de datos se utiliza únicamente el campo IR el cual contiene todas las HRIRs, cuyas dimensiones son 1550x2x256, i.e., existen 256 mediciones de HRIRs para cada una de las 1550 direcciones y para cada oído. Por último, se debe mencionar que todo el procedimiento ya explicado anteriormente, se debe repetir para cada uno de los archivos que se encuentran en la carpeta dtf_nhs. Finalmente, la creación de la base de datos con la información de todos los usuarios, se lo explica en la siguiente sección.

2.1.3. Creación de la base de datos general

El siguiente punto a tratar es sobre la creación de una base de datos que contiene la información de las direcciones y HRIRs de todos los sujetos, las cuales fueron extraídas anteriormente para cada archivo SOFA.

Para llevarlo acabo, lo primero a crear es una matriz que contenga todas las HRIRs de los sujetos para ambos oídos, ya que esta integra la base de datos.

Tabla 2.2: Base de datos definitiva

Sujeto	Azimut	Elevación	Oído	HRTF
1	10	30	1	[1x73]
2	12.5	-30	1	[1x73]
3	30	10	2	[1x73]
4	35	0	1	[1x73]
5	40	80	2	[1x73]
6	42.5	15	2	[1x73]
...	...	...	...	...

2.1.4. Preprocesamiento de las HRIRs

Una vez creada la base de datos general, es necesario preprocesar las HRIRs para convertirlas en HRTFs.

Por lo tanto, se utiliza el Script llamado impulsoafunciondetransferencia.m, el cual aplica la Transformada de Fourier a las respuestas de impulso con el objetivo de obtener las HRTFs.

Segmento de código 2.2: Script impulsoafunciondetransferencia.m

```
N=size(HRIR_todos,2); %obtengo el numero de componentes de frecuencia (256)
%aplico transformada de Fourier para todas las HRIRs del sujeto
%y calculo su modulo ya que solo se necesita las magnitudes
todas_HRTF=abs(fft(HRIR_todos,N,2))/N;
%se toma la mitad de componentes de frecuencia (129) ya que se repiten
%al aplicar la transformada de Fourier
mitad_HRTF=todas_HRTF(:,1:floor(N/2)+1);
%debido a que se recorta la mitad de componentes de frecuencia
%se pierde esa energia por lo que se duplica la energia
%a la mitad de las componentes escogidas
mitad_HRTF(:,2:end-1)=2*mitad_HRTF(:,2:end-1);
%esta es la HRTF que se obtiene finalmente
HRTF=mitad_HRTF;
```

Luego de haber cambiado las HRIRs por HRTFs, se procede a hacer un último preprocesamiento que se aplica sobre las HRTFs, el cual consiste en filtrar las 129 componentes de frecuencia a través de un filtro ideal pasa banda cuya banda de frecuencia va desde los 200 Hz hasta los 14 KHz [35], con lo cual se preserva 73 componentes.

Finalmente, trás realizar todos los procedimientos descritos anteriormente, se tiene la base de datos general descrita en la Tabla 2.2, la cual se utiliza para aplicar en los cuatro métodos de interpolación que se describen en las siguientes Secciones.

2.2. Algoritmos de interpolación para HRTFs

La presente Sección está destinada a indicar los algoritmos implementados para cada uno de los cuatro métodos de interpolación. Un parámetro importante en el proceso de interpolación son los pasos de azimut y elevación, los cuales sirven para declarar puntos vecinos de todo el conjunto de puntos (ubicaciones de las fuentes sonoras). En otras palabras, los pasos indican cada cuanto saltos de puntos a un punto se lo asigna como punto vecino. De

manera que al referirse a los pasos de azimut, quiere decir que se asignan puntos vecinos cada cierto número de salto de puntos que se encuentran a lo largo de una determinada elevación. A continuación en la Figura 2.4 se presenta un ejemplo de la asignación de vecinos y objetivos dado un paso de azimut igual a 2.

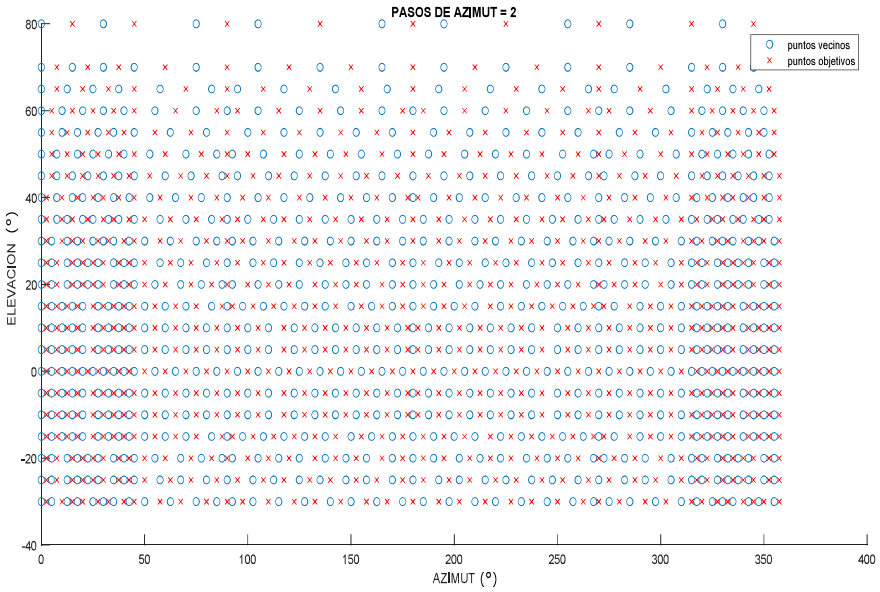

Figura 2.4: Ejemplo pasos de azimut.

Por otro lado para los pasos de elevación, estos se encargan de asignar puntos vecinos de un conjunto de puntos que se encuentran a lo largo de un determinado azimut. Un ejemplo ilustrativo es el de la Figura 2.5, en el cual se asignan vecinos y objetivos dado un paso de elevación igual a 2.

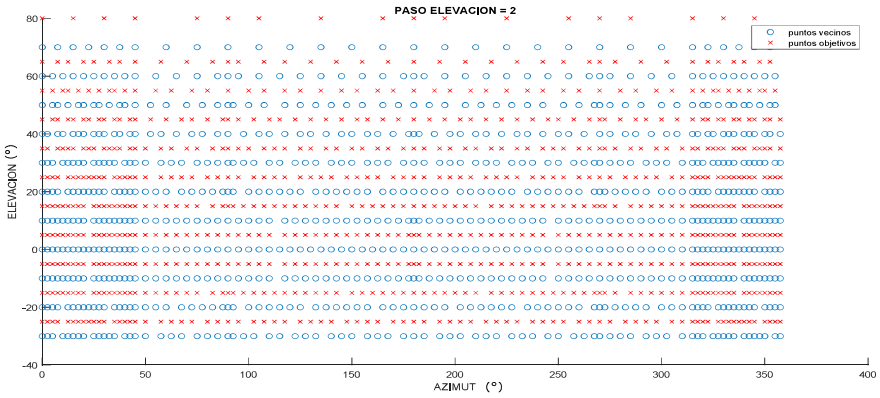

Figura 2.5: Ejemplo pasos de elevación.

Por último, también existe la posibilidad de que en un mismo conjunto de puntos, exista el uso de ambos pasos (i.e. azimut y elevación), los cuales se aplican en los métodos de interpolación bilineal, triangular esférica y splines cúbicos. mientras que el método de inter-

polación lineal adyacente utiliza solamente el paso de azimut ó el de elevación, más no las dos al mismo tiempo. De igual forma, la Figura 2.6 es un ejemplo ilustrativo en donde se aplica la combinación de pasos de azimut y elevación iguales a 2 para la declaración de vecinos y objetivos.

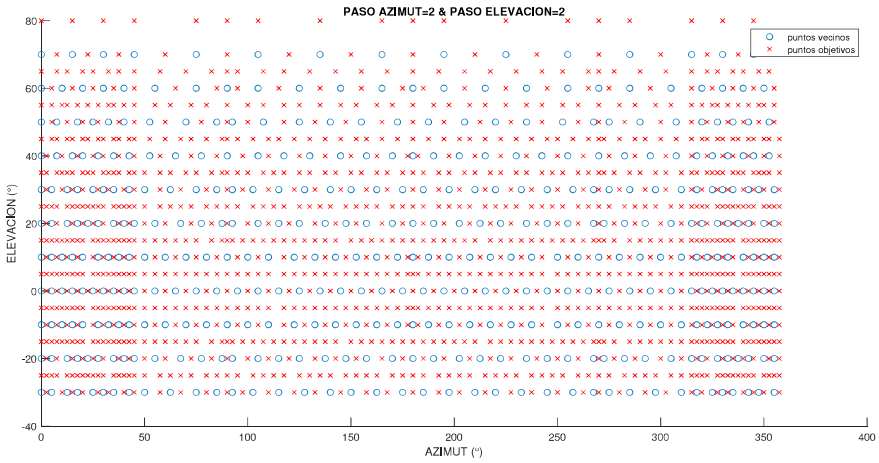

Figura 2.6: Ejemplo combinación de ambos pasos.

2.2.1. Algoritmo para la interpolación lineal adyacente

Este tipo de interpolación tiene dos variantes que son lineal por azimut y lineal por elevación. A continuación se describen ambas formas.

2.2.1.1. Algoritmo de interpolación lineal por azimut

Este método es uno de los más sencillos de implementar, debido a que su interpolación se da únicamente a nivel de azimut, i.e., en los puntos cuyos azimuts se encuentren en una determinada elevación. Lo primero a realizar es la asignación de puntos vecinos (i.e. HRTFs conocidas) y objetivos (HRTFs por interpolar) dado un determinado paso de azimut. Luego se procede a calcular los pesos con las Ecuaciones 1.10 y 1.11 para cada uno de los vecinos de un determinado objetivo. El siguiente segmento de código extraído de la Función lineal_adyacente_azimut.m muestra la implementación en Matlab de este método.

Segmento de código 2.3: Función lineal_adyacente_azimut.m

```
%este lazo se encarga de calcular los pesos de los 2 vecinos
%para cada uno de los puntos objetivos
for i=1:length(tab_ang_veci_infe_izq.azimut)
    peso_superior(i)=(obje_orden_der.azimut(i)...
        -tab_ang_veci_infe_der.azimut(i))...
        /(tab_ang_veci_supe_der.azimut(i)-tab_ang_veci_infe_der.azimut(i));
    peso_inferior(i)=1-peso_superior(i);
end
peso_superior=peso_superior';
peso_inferior=peso_inferior';
```

Finalmente, ya calculado todos los pesos de los vecinos para cada punto objetivo, se procede a calcular las HRTFs aproximadas con la Ecuación 1.12, la cual se encuentra implementada en la Función interpolacion_lineal_adyacente_azimut.m.

Segmento de código 2.4: Función interpolacion_lineal_adyacente_azimut.m

```
for i=1:length(tab_ang_veci_infe_izq.azimut)
    HRTF_INTER_IZQ(i,1:length(OBJETIVO_IZQ.HRTF(1,:)))=(peso_superior(i)...
        *tab_ang_veci_supe_izq.HRTF(i,:))+(peso_inferior(i)...
        *tab_ang_veci_infe_izq.HRTF(i,:));
    HRTF_INTER_DER(i,1:length(OBJETIVO_IZQ.HRTF(1,:)))=(peso_superior(i)...
        *tab_ang_veci_supe_der.HRTF(i,:))+(peso_inferior(i)...
        *tab_ang_veci_infe_der.HRTF(i,:));
end
OBJETIVO_IZQ.HRTF_INTER=HRTF_INTER_IZQ;
OBJETIVO_DER.HRTF_INTER=HRTF_INTER_DER;
OBJETIVO_FINAL=[OBJETIVO_IZQ;OBJETIVO_DER];
```

2.2.1.2. Algoritmo de interpolación lineal por elevación

Para esta interpolación, el algoritmo es semejante al de la lineal adyacente por azimut, solo que la interpolación esta vez es a nivel de elevación, i.e, se da en los puntos cuyas elevaciones se encuentren en un determinado azimut. Sin embargo, existe una inconsistencia al momento de declarar vecinos y objetivos de acuerdo al paso de elevación, debido a que el método de interpolación lineal adyacente asume que la malla en la que las HRTFs fueron medidas es uniforme. Para el caso de la base de datos ARI esto no se cumple ya que fue medida en una malla no uniforme, lo que nos obliga a adaptar el método para resolver este problema. Un ejemplo de este inconveniente se aprecia en la Figura 2.7, en la cual se puede observar que existen puntos objetivos con vecinos bien definidos y no definidos debido al problema ya mencionado de no uniformidad.
Por tal motivo, se utiliza el concepto de distancias euclidianas, las cuales consisten en calcular la distancia entre dos puntos en coordenadas cartesianas. La Ecuación 2.1 es la que indica como calcular dicha distancia.

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2.1}$$

donde
d_E : Es la distancia euclidiana.
P_1, P_2 : Son puntos en coordenadas cartesianas entre los que se obtendrá la distancia.
(x_1, y_1) : Son las coordenadas cartesianas de P_1.
(x_2, y_2) : Son las coordenadas cartesianas de P_2.

De esta manera, cuando se calculen las distancias entre todos los puntos vecinos de una

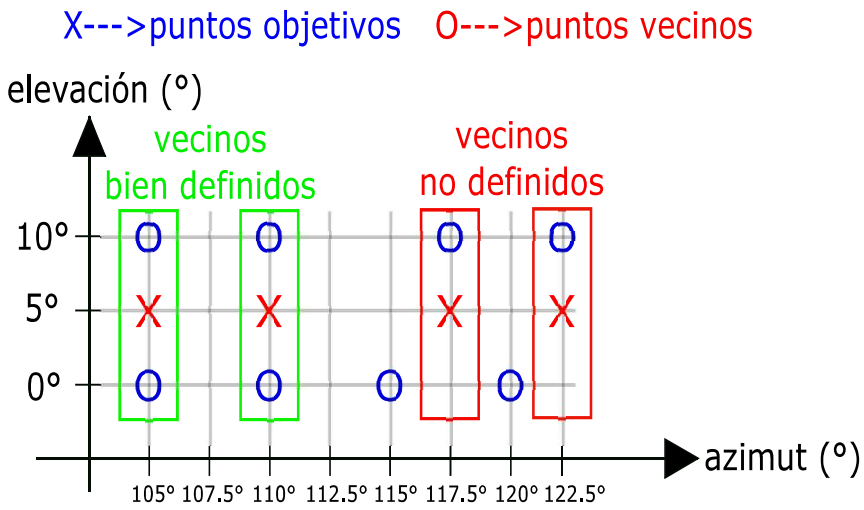

Figura 2.7: Problema de no uniformidad en elevación de la base de datos ARI.

determinada elevación con respecto al objetivo, se determina el punto vecino cuya distancia euclidiana sea la más corta entre todos los vecinos en esa elevación. La figura 2.8 muestra un ejemplo ilustrativo de el procedimiento de selección de vecino para el caso no uniforme.

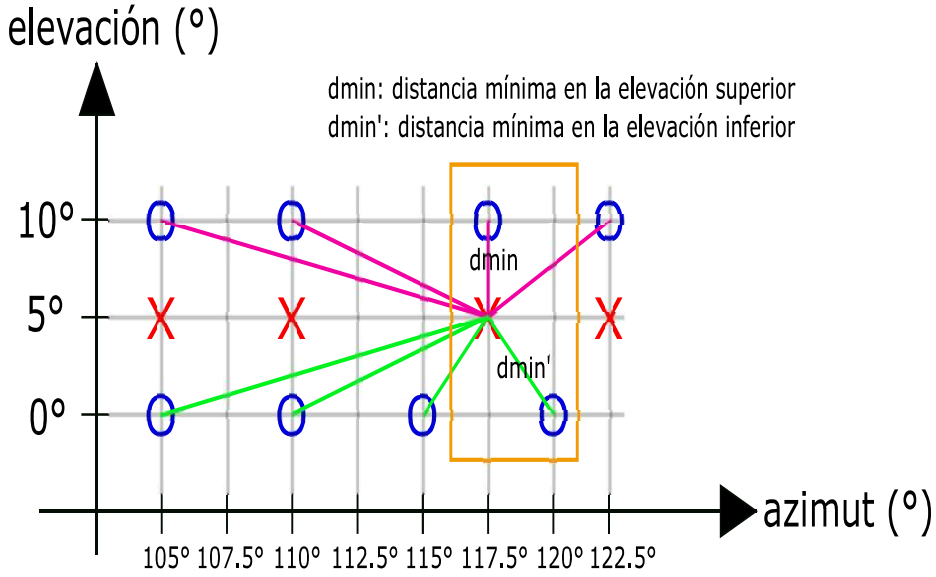

Figura 2.8: Aplicación de la distancia euclidiana como solución a la no uniformidad.

El procedimiento mencionado, se implementó en Matlab como una función cuyo nombre es euclidiano.m, el cual se encarga de buscar los dos vecinos más cercanos al objetivo cuando se interpole por elevación considerando la no uniformidad de la malla en que las HRTFs fueron medidas. Pero previamente se debe convertir las coordenadas esféricas a cartesianas, en donde la función de Matlab sph2cart se encarga de dicha tarea.

Segmento de código 2.5: Función euclidiano.m

```
function [Y]=euclidiano(obje,veci)
for i=1:size(obje,1)
        for j=1:size(veci,1)
            %pdist2 es la que calcula la distancia euclidiana
            distancia(j)=pdist2(obje(i,:),veci(j,:));
        end
        %identifica la distancia mas corta y su ubicacion
        [valor,idxvalor]=min(distancia);
        ubicacion(i)=idxvalor;
end
Y=ubicacion;
```

Luego de aplicar la solución a dichas no uniformidades para poder asignar los vecinos y objetivos, se procede al cálculo de los pesos de los 2 vecinos para cada objetivo, con las Ecuaciones 1.13 y 1.14. El segmento de código 2.6 extraído de la Función lineal_adya ... cente_elevacion2.m calcula los pesos de los vecinos.

Segmento de código 2.6: Función lineal_adyacente_elevacion2.m

```
%en este lazo se calcula los pesos de cada punto objetivo
%con sus respectivos 2 vecinos
for i=1:length(tabla_objetivo_gen_izq.elevacion)
    peso_superior(i)=(tabla_objetivo_gen_izq.elevacion(i)-...
        tabla_vecino_infe_gen_izq.elevacion(i))/...
        (tabla_vecino_supe_gen_izq.elevacion(i)-...
        tabla_vecino_infe_gen_izq.elevacion(i));
    peso_inferior(i)=1-peso_superior(i);
end
peso_superior=peso_superior';
peso_inferior=peso_inferior';
```

Para concluir con el método de interpolación, se utiliza la Ecuación 1.15 para implementarlo en la Función interpolacion_lineal_adyacente_elevacion.m el cual calcula las HRTFs interpoladas o aproximadas en los puntos objetivos.

Segmento de código 2.7: Función interpolacion_lineal_adyacente_elevacion.m

```
%este lazo calcula la HRTF aproximada para cada objetivo
%en donde usa las HRTFs de sus 2 vecinos con sus respectivos pesos
for i=1:length(OBJETIVO_IZQ.elevacion)
    HRTF_INTER_IZQ(i,1:length(OBJETIVO_IZQ.HRTF(1,:)))=...
        (peso_superior(i)*tabla_vecino_supe_gen_izq.HRTF(i,:))+...
        (peso_inferior(i)*tabla_vecino_infe_gen_izq.HRTF(i,:));
    HRTF_INTER_DER(i,1:length(OBJETIVO_IZQ.HRTF(1,:)))=...
        (peso_superior(i)*tabla_vecino_supe_gen_der.HRTF(i,:))+...
        (peso_inferior(i)*tabla_vecino_infe_gen_der.HRTF(i,:));
end
```

2.2.2. Algoritmo para la interpolación bilineal

El presente método se diferencia de los anteriores por la manera de buscar vecinos para cada uno de sus puntos objetivos de la malla de medición de HRTFs. Esta interpolación comienza con la asignación de vecinos y objetivos a través del uso de los pasos de azimut y elevación al mismo tiempo. Después de esto, cada punto objetivo debe buscar los 4 puntos vecinos más cercanos, de manera que se forme una cuadrícula de mediciones. Un ejemplo ilustrativo es el de la Figura 2.9 que muestra como los cuatro puntos vecinos forman una cuadrícula de mediciones con pasos de azimut y elevación igual a 2.

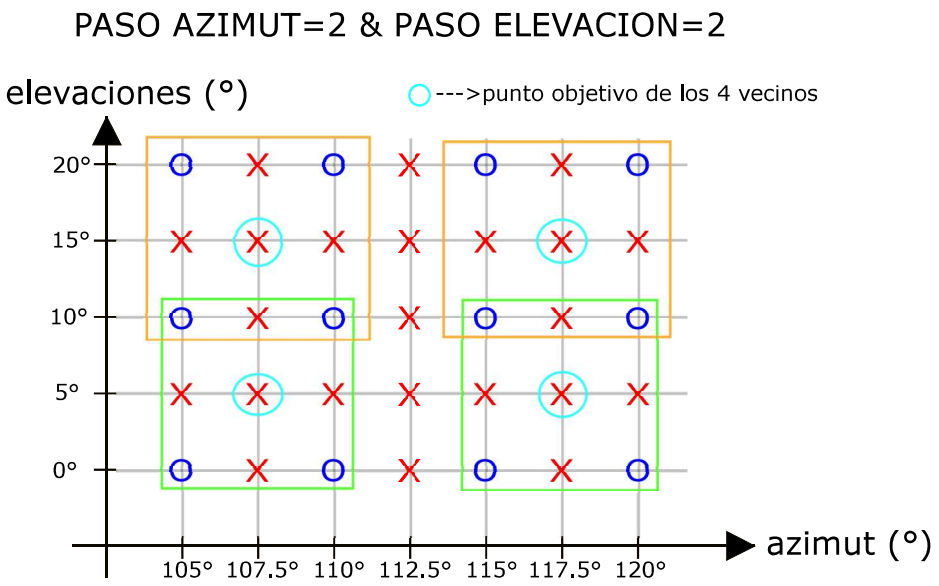

Figura 2.9: Interpolación Bilineal en malla uniforme de mediciones de HRTFs.

Para lo que corresponde a la búsqueda de los 4 vecinos más cercanos con respecto al objetivo, el Segmento de código 2.8 que se obtiene de la Función bilineal.m, nos permite dicha tarea.

Segmento de código 2.8: Función bilineal.m

```
%este lazo se encarga de buscar los 4 vecinos mas cercanos
%para cada uno de los puntos objetivos que existen
for i=1:size(OBJ,1)
    %correction_four_points es una funcion implementada que busca 4 vecinos
    %cercanos al objetivo
    four_neighbours=[four_neighbours;correction_four_points(OBJ(i,:),...
    VEC_modify,elev_unicas1,elev_unicas2)];
end
```

Sin embargo, la interpolación se da correctamente siempre y cuando la malla de medición sea uniforme tal como se vio en la Figura 2.9. De manera que si la malla de medición es no uniforme, no podemos utilizar directamente las Ecuaciones 1.16 para calcular los pesos de dichos vecinos. En otras palabras, cuando la malla es no uniforme, al buscar los 4 vecinos

más cercanos no necesariamente se forma un cuadrado, sino un polígono irregular de 4 lados, tal como se aprecia en la Figura 2.10 para una malla de mediciones no uniforme con vecinos y objetivos para pasos de azimut y elevación igual a 2.

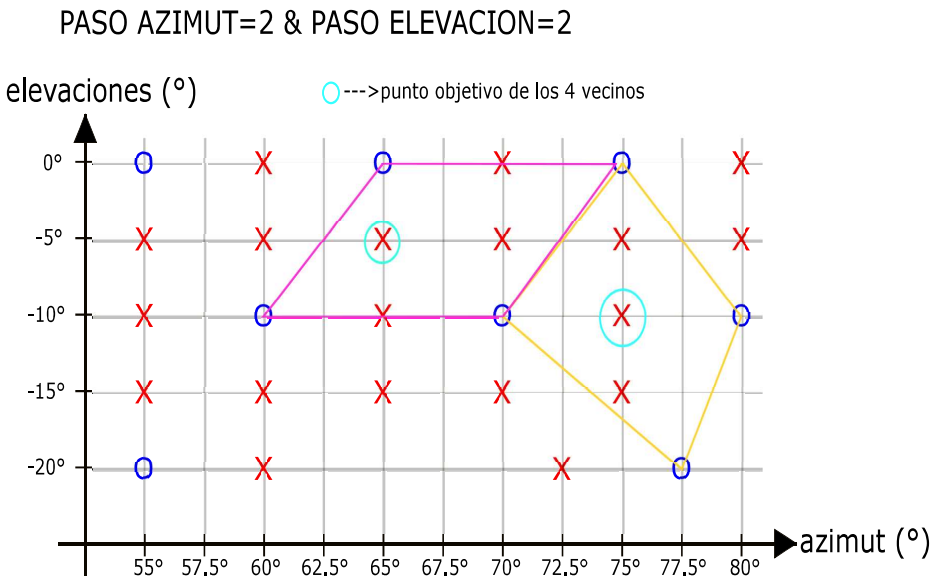

Figura 2.10: Interpolación Bilineal en malla de mediciones HRTF no uniforme.

Con respecto a las ecuaciones de los pesos para el caso de ser una malla de mediciones no uniforme, es claro que la Ecuación 1.16 no tiene validez para aquello. Sin embargo, como solución ante esta irregularidad, se realiza un análisis matemático sobre el origen de la ecuación mencionada anteriormente.

El origen de las fórmulas que permiten determinar los pesos en el método bilineal, nacen de una razón entre áreas que se producen al momento de dividir a la cuadrícula de medición en 4 partes a partir de la ubicación interior del punto objetivo. En otras palabras, cada peso de los 4 vecinos no es más que una razón entre su proporción de área y el área total de la cuadrícula de medición, tal como se ve en la Figura 2.11.

Además es importante mencionar que el área más grande de las 4, corresponde al punto vecino más cercano al objetivo, mientras que el área más pequeña corresponde al punto vecino más lejano al objetivo.

De manera que al analizar las áreas 1, 2, 3 y 4 en términos de azimuts y elevaciones resulta las siguientes ecuaciones:

$$\begin{aligned} area1 &= \Delta\theta * (\phi_g - \Delta\phi) \\ area2 &= \Delta\theta * \Delta\phi \\ area3 &= \Delta\phi * (\theta_g - \Delta\theta) \\ area4 &= (\phi_g - \Delta\phi) * (\theta_g - \Delta\theta) \end{aligned} \tag{2.2}$$

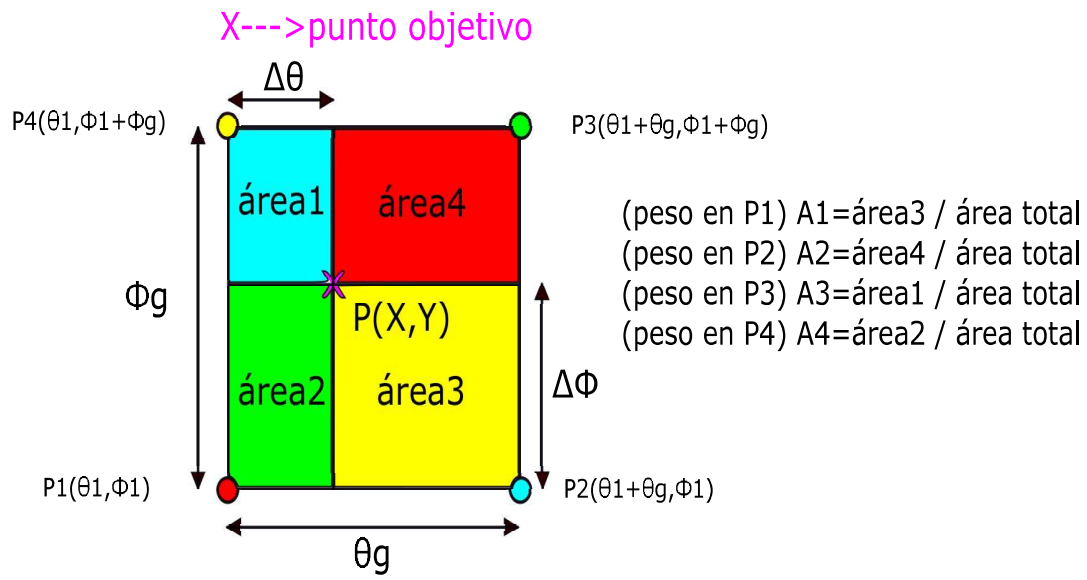

Figura 2.11: Origen del cálculo de pesos en interpolación bilineal.

Por tanto, los pesos de los cuatro vecinos quedan de la siguiente forma:

$$
\begin{aligned}
A_1 &= \frac{\Delta\theta * (\phi_g - \Delta\phi)}{\theta_g * \phi_g} \\
A_2 &= \frac{\Delta\theta * \Delta\phi}{\theta_g * \phi_g} \\
A_3 &= \frac{\Delta\phi * (\theta_g - \Delta\theta)}{\phi_g * \theta_g} \\
A_4 &= \frac{(\phi_g - \Delta\phi) * (\theta_g - \Delta\theta)}{\phi_g * \theta_g}
\end{aligned}
\tag{2.3}
$$

Observe que si se define

$$
A_\theta = \frac{\Delta\theta}{\theta_g} \quad A_\phi = \frac{\Delta\phi}{\phi_g}, \tag{2.4}
$$

se llegan a las mismas expresiones explicadas en las Ecuaciones 1.16. Por tanto, así queda demostrado el análisis matemático que describe el origen de las ecuaciones para el cálculo de pesos en el método Bilineal.
Entonces, considerando el anterior análisis y para adaptarlo a la no uniformidad de la base de datos, la Figura 2.12 indica como combinar el concepto de razón de áreas en la malla de medición no uniforme. Por tal motivo, el presente proyecto propone trazar una cruz en el punto objetivo, de manera que se divida a los polígonos de 4 lados en 4 partes (i.e. 4 áreas).

Después se determina sus áreas para finalmente obtener los pesos de cada uno de los 4 vecinos. Sin embargo, vuelve a nacer un nuevo problema. Todos los puntos pertenecientes a la malla de mediciones de HRTFs, sean objetivos o vecinos son parte de una esfera donde

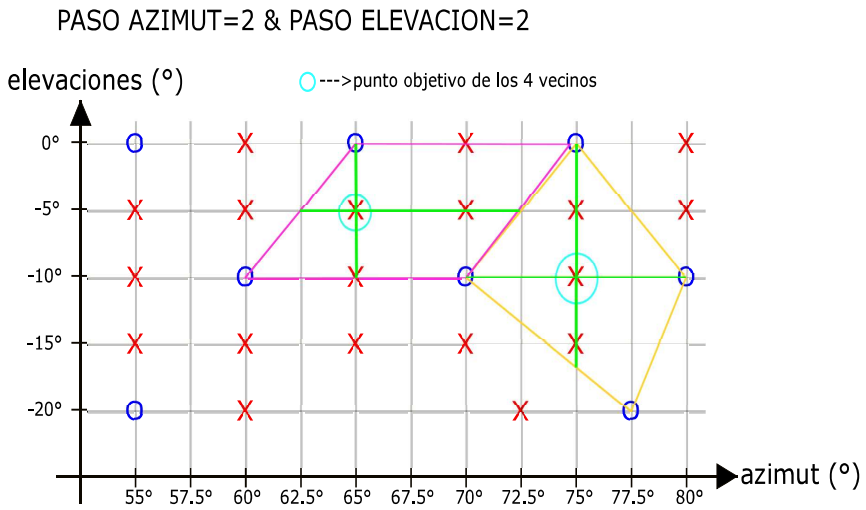

Figura 2.12: División en 4 áreas en malla de mediciones HRTFs no uniforme.

obviamente sus coordenadas son 3D. Para resolver esta inconsistencia, se mapea los puntos 3D a 2D donde se asume que estos puntos pertenecen a una variedad (*Manifold* [36]), i.e., asúmase que los puntos son localmente lineales y globalmente no lineales (i.e. porque forman una esfera). Un ejemplo ilustrativo es el que se presenta en la Figura 2.13, donde se aprecia que de manera global la tierra es una esfera (i.e. no lineal) pero si se ubica en un punto específico sobre la esfera, el entorno se puede aproximar como una superficie 2D lineal. Observe que las mediciones de las HRTFs son realizadas sobre una esfera por lo que asumir que puntos cercanos forman una variedad es una suposición adecuada. Este mapeo

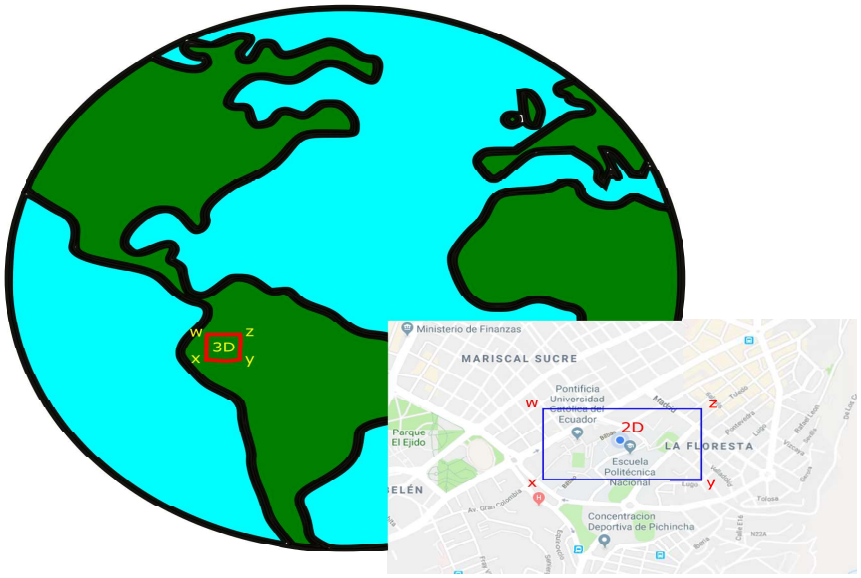

Figura 2.13: Diferencia entre puntos globales no lineales (i.e. puntos 3D) y puntos linealmente locales (i.e. puntos 2D).

puede ser realizado con PCA (*Principal Component Analysis* [37]), el cual es un conocido y ampliamente utilizado método de reducción de dimensionalidad que permite mapear los

puntos 3d a un plano 2d. Por tanto, hacemos uso de la función nativa de Matlab, llamada pca, la cual se encuentra implementada en la Función pesos_bilineal.m.
Finalmente, ya realizado el cambio de dimensionalidad, se procede a dividir al polígono de medición en 4 partes para posteriormente calcular las áreas y después sus pesos.

2.2.3. Algoritmo para la interpolación triangular esférica

El presente método de interpolación es idéntico al de la Bilineal, solo que este tiene un proceso adicional al momento de seleccionar sus vecinos. Es decir, al momento de encontrar sus 4 puntos vecinos más cercanos, procede a determinar cuál de los puntos es el más cercano con respecto al punto objetivo para luego escoger los dos puntos adyacentes con respecto al punto más cercano. En resumen, cada punto objetivo únicamente tendrá 3 puntos vecinos, los cuales forman un triángulo de mediciones que contienen al punto objetivo, tal como se muestra en la Figura 2.14 para una malla de mediciones con pasos de azimut y elevación igual a 2.

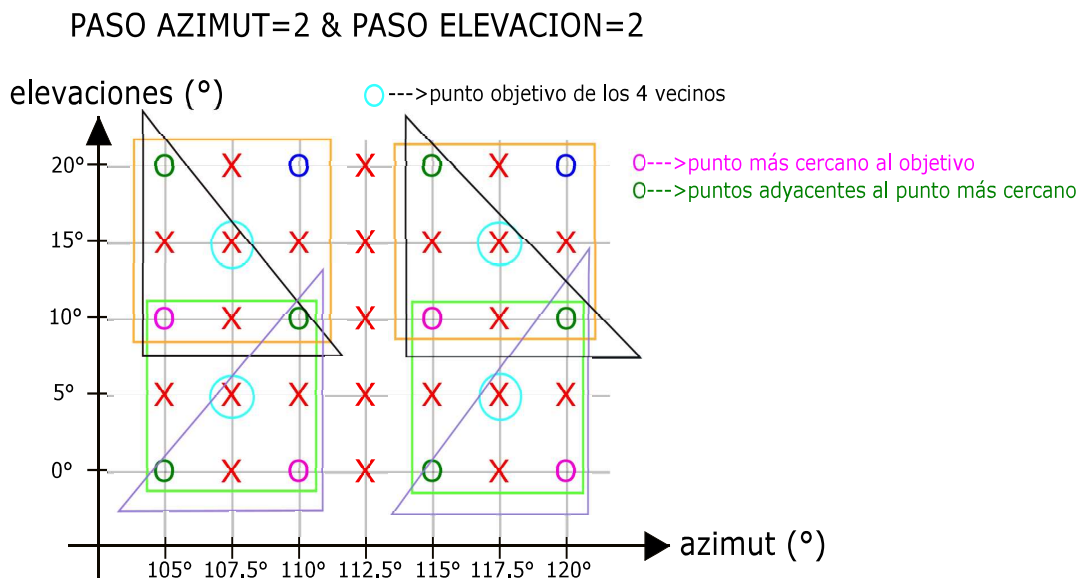

Figura 2.14: Interpolación Triangular Esférica en malla de mediciones HRTF uniforme.

Para lo que corresponde a la búsqueda de los 4 vecinos más cercanos con respecto al objetivo, de igual manera, la Función bilineal.m se encarga de esa acción. Luego implementamos una función llamada triang_spherical.m, la cual determina los 3 puntos vecinos más cercanos al objetivo.
Sin embargo, esta interpolación se da correctamente siempre y cuando la malla de medición sea uniforme tal como se vio en la Figura 2.14. De manera que si la malla de medición es no uniforme, no podemos utilizar directamente las Ecuaciones 1.18 para calcular los pesos de dichos vecinos. En otras palabras, cuando la malla es no uniforme, al buscar los 3 vecinos más cercanos no necesariamente se forma un triángulo rectángulo, tal como se aprecia en la Figura 2.15 para una malla de mediciónes no uniforme con vecinos y objetivos para pasos de azimut y elevación igual a 2.

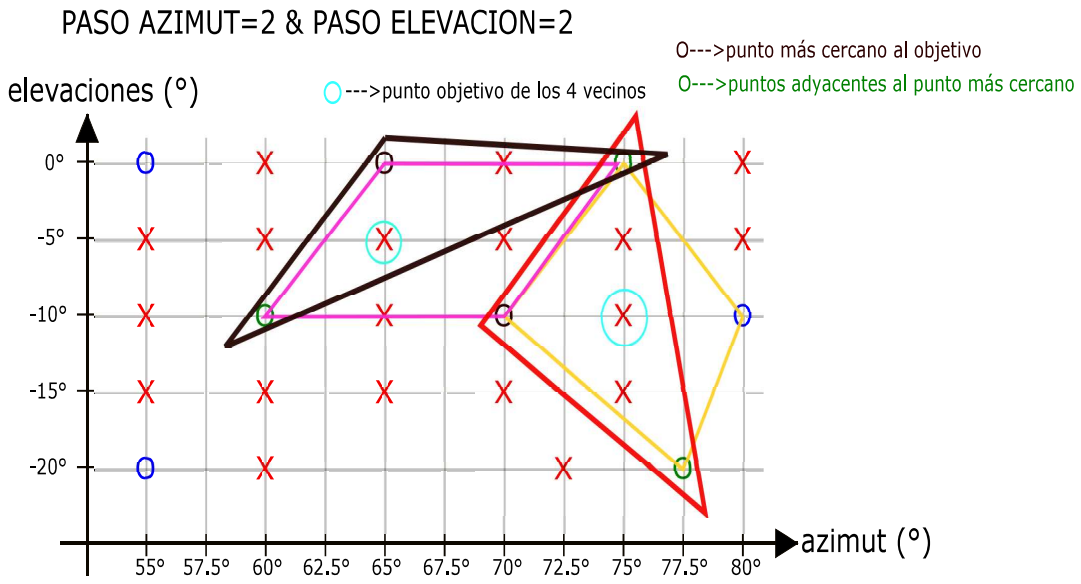

Figura 2.15: Interpolación Triang. Esf. en malla de mediciones HRTF no uniforme.

Por tal motivo, al igual que en la interpolación Bilineal, se analiza matemáticamente las ecuaciones ya mencionadas, cuyo origen es semejante al método anterior.

En cuanto al origen de las fórmulas relacionadas con la interpolación triangular esférica, estas nacen de una razón entre áreas que se producen al momento de dividir al triángulo de mediciones en 3 partes (3 triángulos) a partir de la unión de 3 líneas trazadas desde los vértices del triangulo hacia la ubicación interior del punto objetivo. En otras palabras, cada peso de los 3 vecinos no es más que una razón entre su proporción de área y el área total del triángulo de medición, tal como se ve en la Figura 2.16.

Además es importante mencionar que cada una de las 3 áreas corresponde al punto vecino opuesto, i.e., con el punto que no tenga adyacencia a dichas áreas.

De manera que al analizar las áreas 1, 2 y 4 en términos de azimuts y elevaciones resulta las siguientes ecuaciones:

$$\begin{aligned} AREA1 &= \frac{\theta_g * (\phi_g - \Delta\phi) - (\Delta\theta * \phi_g)}{2} \\ AREA2 &= \frac{\Delta\theta * \phi_g}{2} \\ AREA4 &= \frac{\Delta\phi * \theta_g}{2} \end{aligned} \tag{2.5}$$

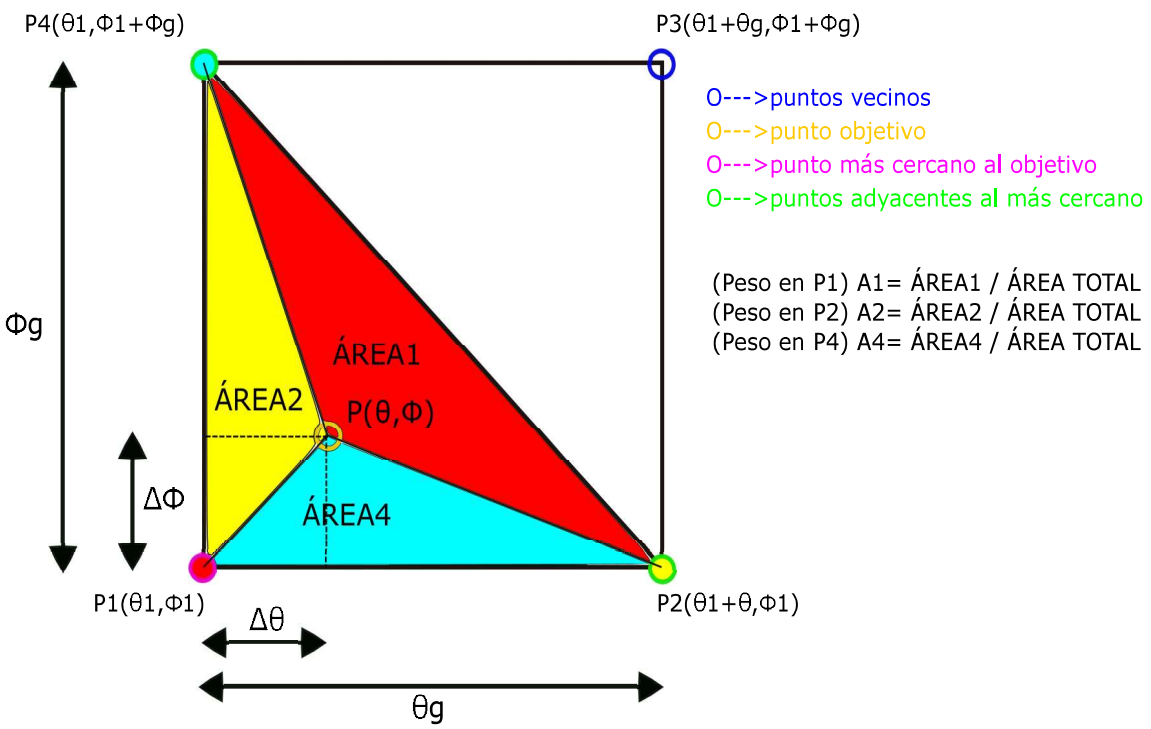

Figura 2.16: Origen del cálculo de pesos en la interpolación triangular esférica.

Por tanto, los pesos de los tres vecinos quedan de la siguiente forma:

$$
\begin{aligned}
A_1 &= \frac{\frac{\theta_g*(\phi_g-\Delta\phi)-(\Delta\theta*\phi_g)}{2}}{\frac{\phi_g*\theta_g}{2}} = 1-\frac{\Delta\phi}{\phi_g}-\frac{\Delta\theta}{\theta_g} \\
A_2 &= \frac{\frac{\Delta\theta*\phi_g}{2}}{\frac{\phi_g*\theta_g}{2}} = \frac{\Delta\theta}{\theta_g} \\
A_4 &= \frac{\frac{\Delta\phi*\theta_g}{2}}{\frac{\phi_g*\theta_g}{2}} = \frac{\Delta\phi}{\phi_g}
\end{aligned}
\tag{2.6}
$$

Observe que al definir de igual forma que en el método de interpolación Bilineal

$$
A_\theta = \frac{\Delta\theta}{\theta_g} \quad A_\phi = \frac{\Delta\phi}{\phi_g}, \tag{2.7}
$$

se llegan a las mismas expresiones explicadas en las Ecuaciones 1.18. Por tanto, así queda demostrado el análisis matemático que describe el origen de las ecuaciones para el cálculo de pesos en el método triangular esférico. De igual forma, el método tiene una inconsistencia con respecto a sus coordenadas en 3D, debido a que fueron obtenidos de una esfera y la presente interpolación requiere que sea en 2D. Por tanto también se emplea el uso de pca para su respectivo cambio de dimensionalidad, tal como se aplicó en el método bilineal.
Después de realizar la última modificación a las coordenadas, se procede a determinar las áreas para luego calcular los respectivos pesos de los 3 vecinos que corresponden a cada

punto objetivo. De estas tareas se encarga una función ya implementada en Matlab denomidada triang_spherical.m.
Finalmente, se procede a calcular las HRTF aproximadas o interpoladas con la Ecuación 1.19, la cual está implementada en la Función interpolacion_triang_es ferica.m.

2.2.4. Algoritmo para la interpolación por splines cúbicos

Con respecto a este último método de interpolación, es uno de los más fáciles de implementarse debido a que Matlab contiene una función llamada griddata, la cual tiene la opción de aplicar interpolaciones por splines cúbicos, de manera que dicha función nativa de Matlab se halla implementada en la Función cubic_spline.m.

3. RESULTADOS Y DISCUSIÓN

El presente capítulo trata sobre el análisis de los resultados obtenidos de los cuatro métodos de interpolación en HRTFs a través de gráficas de dispersión y de barras, considerando a su vez la métrica de distorsión espectral y los parámetros que son: suavizado espectral, pasos de azimut-elevación (*puntos vecinos y objetivos*), sujetos y direcciones. Por tanto se discutirán las gráficas resultantes de acuerdo a la métrica y sus parámetros ya mencionados a través de las siguientes Secciones:
La Sección 3.1 explica los resultados de la distorsión espectral en base a la utilización del suavizado espectral. La Sección 3.2 se refiere a los resultados de la distorsión en base a la variación de los pasos de azimut y elevación, mientras que la Sección 3.3 presenta los niveles de distorsión espectral en sujetos. Por último, la Sección 3.4 indica los resultados de distorsión espectral de acuerdo a las direcciones de las fuentes sonoras.

3.1. Análisis del efecto del suavizado espectral

Para empezar, se analizan las gráficas resultantes considerando el parámetro del suavizado espectral o Smoothing, de manera que se concluye si la distorsión espectral disminuye o aumenta tras aplicar la métrica ya mencionada.
Por otra parte, no es posible comparar los cuatro métodos de interpolación para todos los pasos de azimut y elevación, ya que algunos métodos no utilizan ambos pasos al mismo tiempo ó simplemente no existen determinadas combinaciones de pasos de azimut y elevación en dichos métodos, tales que producen un desbalance al momento de compararlos.
Por ejemplo, en los cuatro métodos no existe alguno que utilice paso de azimut y elevación igual a 1 al mismo tiempo, ya que no existiría puntos objetivos, únicamente habría puntos vecinos sin nada que interpolar. Otro ejemplo consiste en que la interpolación lineal adyacente por azimut, no acepta pasos de elevación mayores a 1 o si fuese la lineal adyacente por elevación, tampoco acepta pasos de azimut mayores a 1. En pocas palabras el método de la interpolación lineal adyacente no puede ser comparado con los demás métodos a excepción de una determinada combinación de pasos de azimut y elevación en donde si exista para los cuatro métodos.
En cuanto a la Figura 3.1, se observa la relación entre los pasos de elevación y distorsión espectral sin aplicar suavizado espectral (barras de colores) y con la utilización de la misma (barras de color verde), donde al aumentar el paso de elevación, las distorsiones espectrales también aumentan. Entonces los resultados son correctos, debido a que un incremento de los pasos de elevación equivale a distanciar los puntos vecinos con respecto a los objetivos. Por lo que las HRTFs interpoladas no tendrán una correcta aproximación con respecto a las HRTFs originales.
Además, considerando el parámetro del suavizado espectral, se observa que los valores de distorsión espectral reducen notablemente en los cuatro métodos de interpolación. También en estos métodos se utilize o no el suavizado espectral, tienden a tener casi los mismos valores de distorsión espectral. Esto en consecuencia de que la presente combinación de

pasos causa que la mayoría de interpolaciones de los cuatro métodos se reduzcan a interpolaciones lineales adyacentes por elevación.

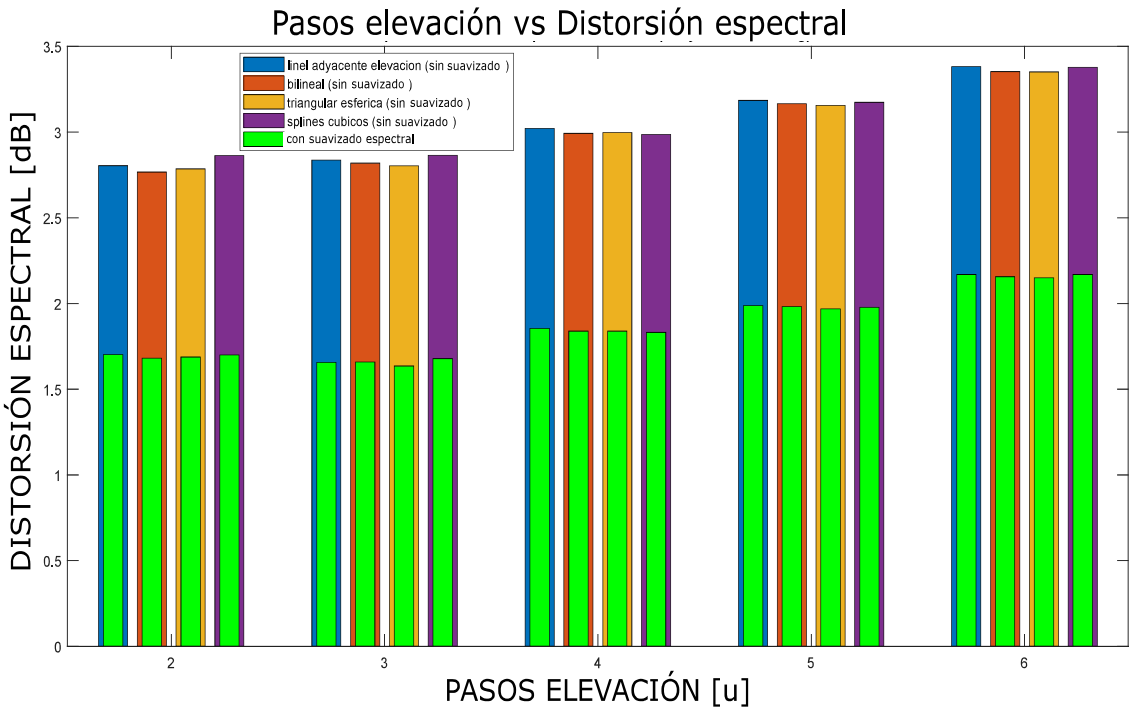

Figura 3.1: Gráfico de barras (Pasos elevación vs Distorsión espectral).

Por otra parte, la Figura 3.2 muestra la relación entre los pasos de azimut y la distorsión espectral cuando se utiliza el suavizado espectral (barras de color verde) y cuando no se lo aplica (barras de colores), de manera que en la figura se observa que al aumentar el paso de azimut, aumenta la distorsión espectral. Esto porque al encontrarse los puntos vecinos más distantes de los objetivos, causan que las HRTFs interpoladas no tengan una correcta aproximación, causando que al determinar la distorsión espectral con las HRTFs interpoladas y originales, produzcan un margen de error grande. En cuanto al suavizado espectral, se observa claramente como en la Figura 3.1, que la utilización de la misma reduce notablemente las distorsiones espectrales.

En cuanto a las distorsiones espectrales en cada uno de los cuatro métodos para diferentes pasos de azimut, se observa que el método que presenta menores distorsiones es la lineal adyacente, debido a que existe un desbalance con respecto al resto de métodos. Es decir, la lineal adyacente contiene únicamente puntos objetivos provenientes de pasos de elevación iguales a 1 y los demás métodos contienen puntos objetivos derivados de pasos de elevación que van desde 1 hasta 6. Entonces, al tener menos puntos objetivos (HRTFs interpoladas), la distorsión espectral no será tan elevada como para el resto de métodos que contienen mayor cantidad de puntos objetivos, los cuales conducen a una distorsión más elevada. Por lo tanto, se descarta momentáneamente dicho método para comparar los tres métodos restantes, de manera que los métodos cuya distorsión espectral es baja, son la bilineal y splines cúbicos. Este resultado se produce al utilizar más puntos vecinos en dichos métodos de interpolación, ya que conducen a menores valores de distorsión espectral, mientras que

el método que presenta mayor distorsión espectral en la mayoría de pasos de azimut, es la triangular esférica, i.e., es el peor método de interpolación.

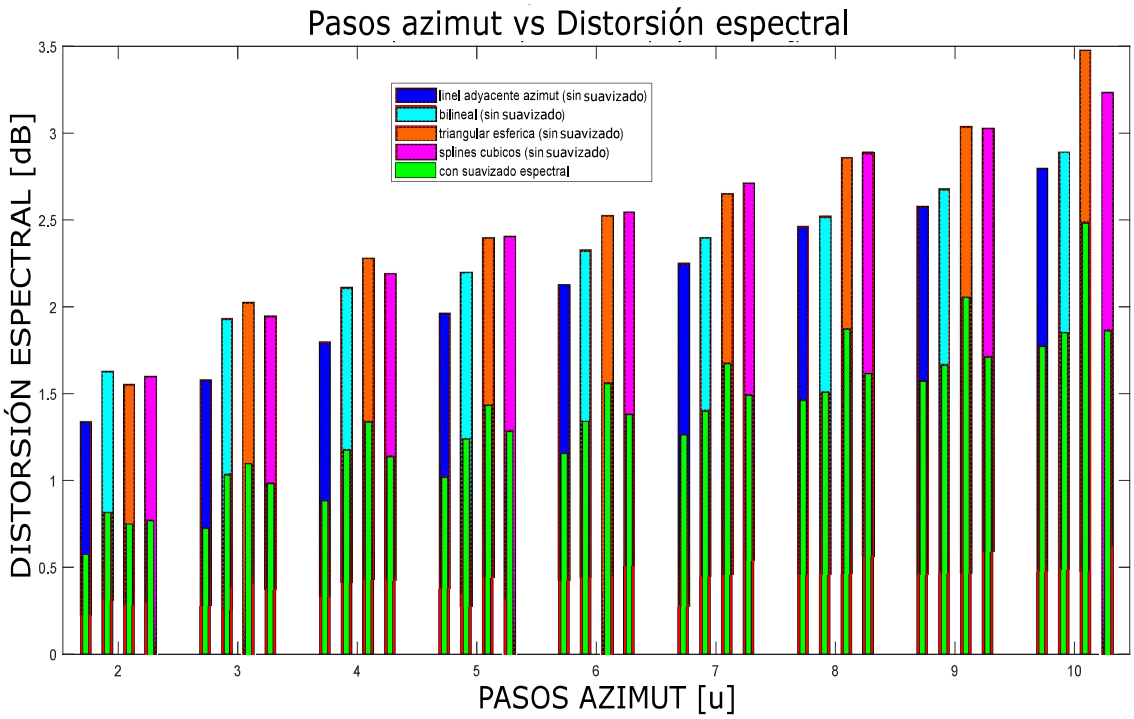

Figura 3.2: Gráfico de barras (Pasos azimut vs Distorsión espectral).

A continuación, la Figura 3.3 muestra mapas de calor que indica la distorsión espectral a través de barras de colores, donde la escala de rojos indica altos niveles de distorsión espectral y la de azules bajos niveles de la misma. Por lo tanto, se puede observar que en los mapas de calor en donde no se utiliza el suavizado espectral, son notables las escalas del rojo conforme se aumente los pasos de azimut o de elevación. En otras palabras, la aplicación de métodos de interpolación sin suavizado espectral, conlleva a altas distorsiones espectrales y aumentarán con el incremento de pasos de azimut o elevación.

Por otro lado, al utilizar el suavizado espectral, se evidencia que en todos los métodos domina la escala del amarillo, y de igual forma, al aumentar los pasos de azimut y elevación, producen elevadas distorsiones espectrales pero no tanto como el caso de no utilizar el suavizado.

Finalmente, después de haber analizado las Figuras 3.1, 3.2 y 3.3, en donde hacen referencia al suavizado espectral, se concluye que es mejor utilizarlo en los cuatro métodos de interpolación, debido a que reduce considerablemente sus valores de distorsión espectral. Además, todo lo mencionado, es corroborado por [1], donde experimentos psicoacústicos demuestran que el suavizado espectral es válido. Luego de esta aclaración, los siguientes resultados utilizan suavizado espectral ya que mejora la interpolación y reduce errores de aproximación.

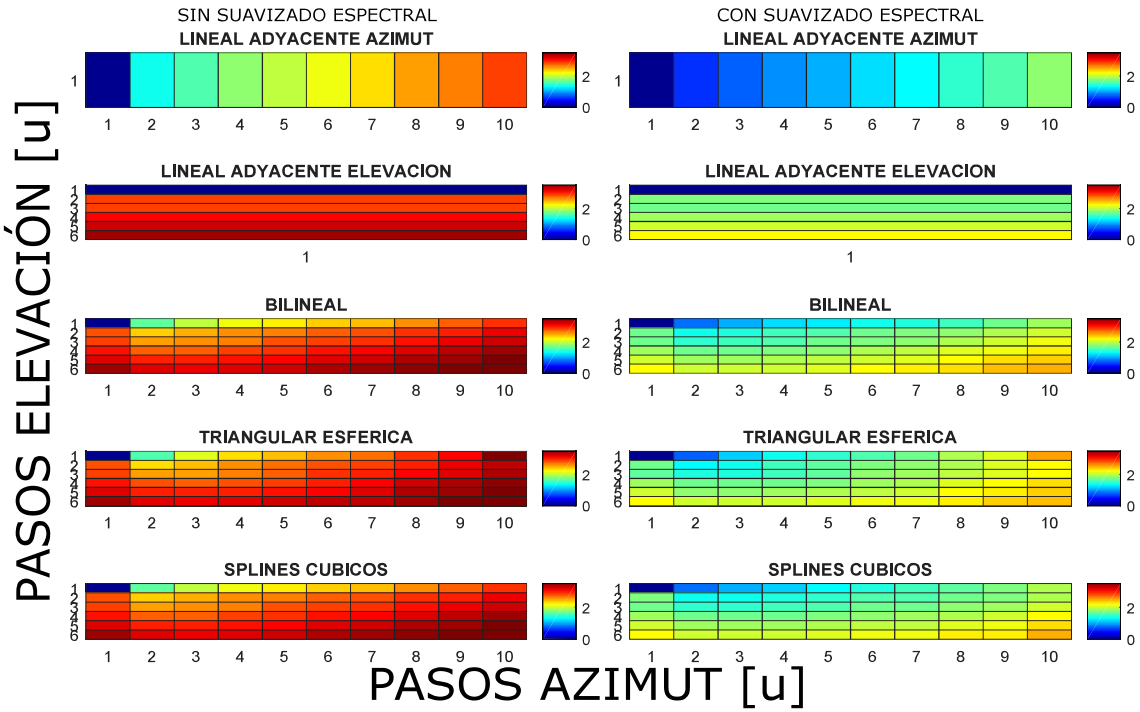

Figura 3.3: Mapa de calor (paso azimut vs paso elevación vs Distorsión espectral (barras de color)).

3.2. Análisis del efecto de los pasos de azimut y elevación

En esta sección, se busca demostrar que las distorsiones espectrales de los métodos de interpolación también dependen bastante de los pasos de azimut y elevación.

Por tal motivo, en la Figura 3.4 se muestra un gráfico de scatter, el cual relaciona los pasos de azimut y elevación con la distorsión espectral media expresada a través de una barra de colores, donde la escala del rojo indica alta distorsión espectral y la escala del azul indica bajos niveles de la misma métrica, considerando únicamente los métodos bilineal, triangular esférica y splines cúbicos. No se considera a la lineal adyacente porque el método exige que uno de los pasos (azimut o elevación) sea igual a 1.

Dicho lo anterior y retomando el análisis de la Figura 3.4, se observa que al incrementar el paso de azimut o el de elevación, la intensidad del rojo va siendo más notorio. En otras palabras, la escala del azul (distorsiones bajas) tiene mayor presencia en los pasos más pequeños. En tanto que la escala del rojo (distorsiones altas) tiene mayor dominio en pasos más elevados. En conclusión, al aumentar los pasos de azimut o elevación, causan incrementos en la distorsión espectral. En otras palabras, al aumentar los pasos, se distancian los puntos vecinos con respecto a los objetivos, de manera que los errores de aproximación van empeorando y estos a su vez conducen a elevados niveles de distorsión espectral.

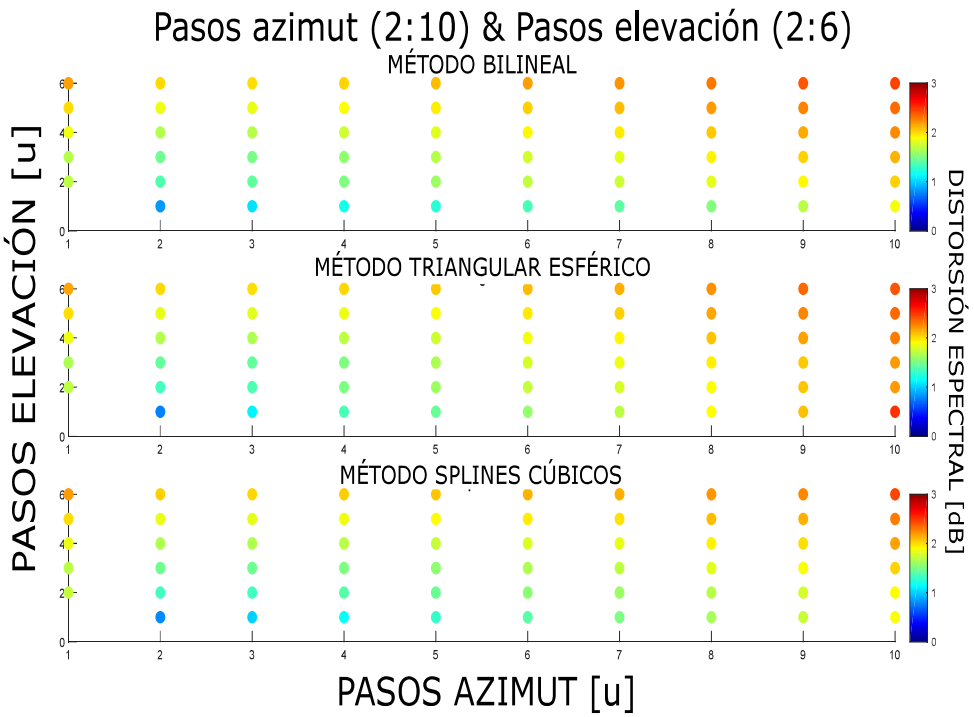

Figura 3.4: Gráfico Scatter (paso azimut vs paso elevación vs Distorsión espectral media) para los 3 métodos, excluyendo al método lineal adyacente.

3.3. Análisis de la distorsión espectral en los sujetos

En esta sección se analizan los niveles de distorsión espectral para cada uno de los métodos de interpolación, en cada uno de los 89 sujetos por intermedio de gráficas de barras y con leyendas, las cuales indican el promedio de distorsión espectral con su respectiva desviación estándar para cada uno de los métodos de interpolación.

Sin embargo, no se pueden comparar los cuatro métodos en todos los pasos, por la inconsistencia explicada anteriormente. Por tal motivo se seleccionan 3 combinaciones de pasos, en donde las 2 primeras combinaciones comparan los 4 métodos y en la última combinación, solo compara 3 métodos, excluyendo a la lineal adyacente.

Entonces, la Figura 3.5 muestra un gráfico de barras para los 89 sujetos y considerando los cuatro métodos para una combinación de pasos de azimut y elevación iguales a 1 y 5 respectivamente, de manera que el método de lineal adyacente que existe junto con los otros 3 métodos es la de elevación. Además, en las distorsiones espectrales se puede observar que sus valores en los cuatro métodos son semejantes y se encuentran cerca de los 2dB, debido a que los métodos tienen distorsiones y desviaciones estándar semejantes. A causa de la presente combinación de pasos, la mayoría de interpolaciones tienden a reducirse en una simple interpolación lineal adyacente.

Por otra parte, tanto la Figura 3.6 como la Figura 3.5, indican las variaciones de distorsión espectral que presentan los cuatro métodos de interpolación en cada sujeto. Considerando la combinación de pasos de azimut y elevación iguales a 8 y 1 respectivamente, de manera que la interpolación lineal adyacente que existe es la de azimut, el cual presenta una me-

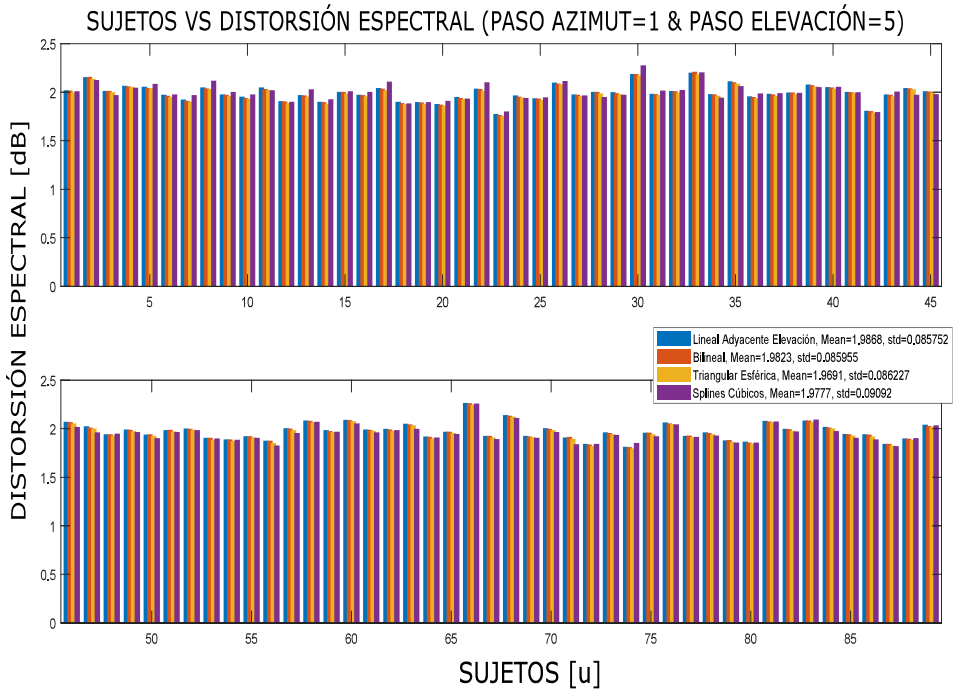

Figura 3.5: Gráfico de barras (Sujetos vs Distorsión espectral) para los 4 métodos.

nor distorsión espectral con respecto a los otros 3 métodos. El resultado se produce por un desbalance en la cantidad de HRTFs interpoladas por cada método, tal como se menciona en el análisis de la Figura 3.2, de manera que la lineal adyacente tenga menor distorsión espectral, mientras que en los otros métodos al tener mayor cantidad de HRTFs interpoladas, tendrán por consiguiente mayor distorsión espectral. Además, existe otra observación, el método que presenta notablemente una mayor distorsión espectral en todos los sujetos, es la triangular esférica. También para esta combinación de pasos se puede ver que las distorsiones en todos los métodos a excepción de la triangular esférica están cerca de los 1.5dB, que en comparación a la Figura 3.5, son menores.

Por último, la Figura 3.7, al igual que en las Figuras 3.5 y 3.6, indican las variaciones de distorsión espectral, pero esta vez para los 3 métodos, donde se excluye a la lineal adyacente, debido a que se utiliza la combinación de pasos de azimut y elevación iguales a 8 y 5 respectivamente.

En cuanto a su análisis, se observa que las distorsiones se encuentran por encima de los 2dB, debido a que la combinación de pasos que se está utilizando, es un caso extremo que se presenta en las interpolaciones. Por lo que al aumentar los pasos, su distorsión también se eleva.

Además, los métodos de interpolación que presenta menores distorsiones, son el de splines cúbicos, seguido de la triangular esférica, ya que el método de splines cúbicos interpola con más de 2 o 3 vecinos, lo que reduce sus niveles de distorsión espectral, mientras que la triangular esférica también muestra una mejor eficiencia , debido a que considera sus 3 puntos vecinos más cercanos para conformar un triángulo de medición para interpolar.

En conclusión y tomando en consideración los análisis relacionados con la distorsión es-

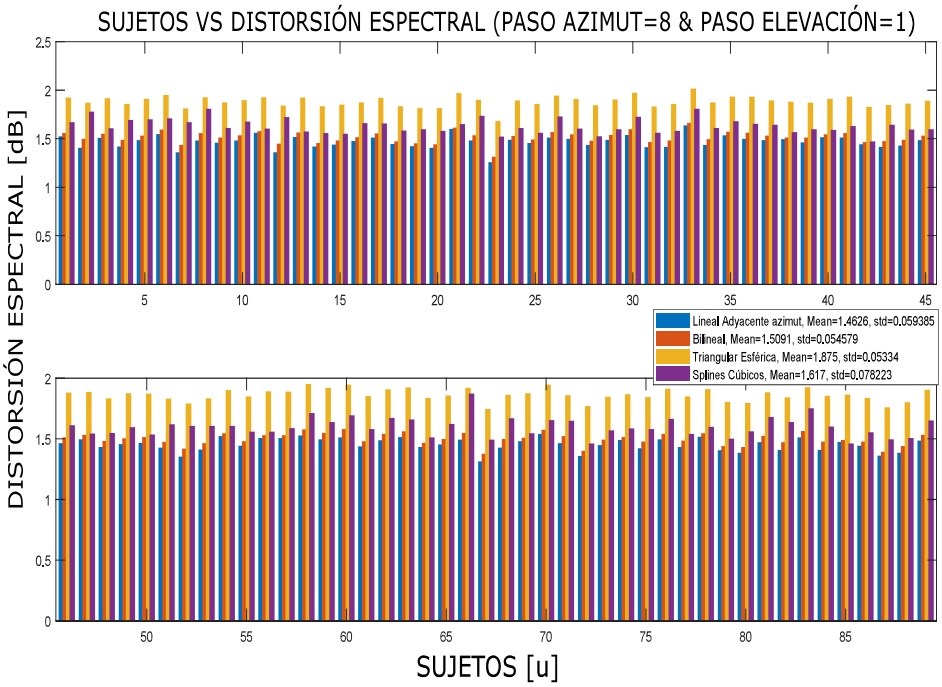

Figura 3.6: Gráfico de barras (Sujetos vs Distorsión espectral) para los 4 métodos.

pectral en esta sección. Los métodos bilineal, triangular esférica y splines cúbicos presentan niveles de distorsión espectral semejantes, debido a que las desviaciones estándar son similares y estas a su vez indican que las diferencias entre las distorsiones de los métodos de interpolación son despreciables.

3.4. Análisis de la distorsión espectral en función de la dirección

En lo que corresponde a esta sección, se continua con el análisis de la distorsión espectral, pero esta vez en base a las direcciones o ubicaciones de las fuentes sonoras, en donde fueron medidas las HRTFs.

Entonces, para llevar a cabo el análisis de la distorsión espectral en la presente sección, se lo realiza en 3 partes: Primero, analizando todas las direcciones de los puntos objetivos generados por los pasos de azimut y elevación en gráficas de scatter. Después, se hace el análisis de dicha métrica por planos (ipsilateral, contralateral y mediano) y por último, se analiza la distorsión espectral en cuatro bandas de frecuencia.

Además, al igual que en análisis anteriores. Este presente análisis también utiliza las 3 combinaciones de pasos, debido a que los cuatro métodos no pueden ser comparados en todas las combinaciones de pasos, de manera que las 2 primeras combinaciones permiten comparaciones entre los 4 métodos de interpolación, mientras que la última combinación solo compara 3 métodos, en donde fue excluida la lineal adyacente.

Aclarado dicho condicionamiento, se procede a analizar la Figura 3.8, la cual muestra 4 gráficas de scatter que relacionan las coordenadas de los puntos objetivos (ángulos de azimut y elevación) con las distorsiones espectrales mediante una barra de colores, cuyos valores

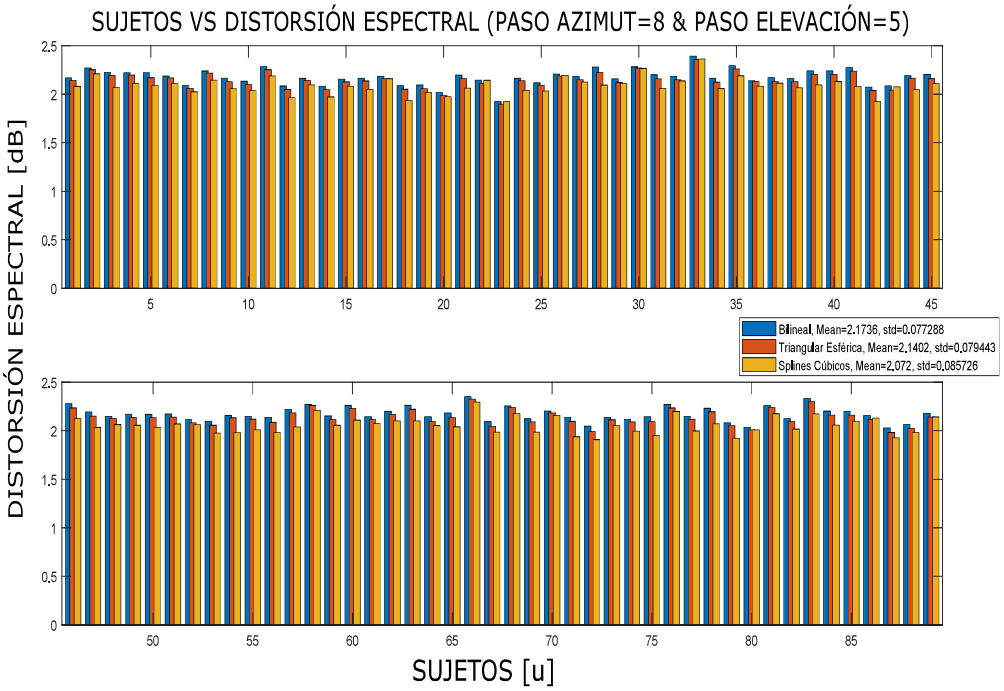

Figura 3.7: Gráfico de barras (Sujeto vs Distorsión espectral) para los 3 métodos.

más bajos son representados por el color azul (0dB). Mientras que los valores más altos de distorsión están representadas por el color rojo (8dB). Por otra parte, las direcciones de los puntos objetivos, fueron generadas a partir de la primera combinación de pasos de azimut y elevación iguales a 1 y 5 respectivamente. En referencia a la barra de colores, se observa que la escala de color azul tiene un dominio total en los cuatro métodos de interpolación, de manera que sus distorsiones no superan más de los 3dB. Sin embargo, existe unas distorsiones espectrales (color celeste) no tan elevadas en determinados rangos de azimut, i.e., entre $70° \leq \theta \leq 110°$ y $250° \leq \theta \leq 290°$. Esto se debe al efecto sombra de la cabeza, para lo cual hay que recordar los sistemas de coordenadas de la Sección 1.4.1.1, en donde para la referencia de un sistema de coordenadas esféricas, considera al lado izquierdo de la cabeza (oído izquierdo) con un azimut (θ) igual a $90°$ y al lado derecho (oído derecho) con un azimut igual a $270°$. Por lo que al momento de medir las HRTFs, si nos encontramos en un determinado oído y a su vez se mide en una dirección donde corresponde al oído contrario, está medición contiene una distorsión elevada, debido a que la cabeza será un obstáculo durante la percepción del sonido (efecto sombra). Es por eso que en dichos rangos de azimuts, al contener los azimuts correspondientes a los oídos, se da una menor intensidad del color azul (distorsión espectral más grande).

Por otra parte, la Figura 3.9 es similar a la Figura 3.8, a diferencia de que esta considera los ángulos de azimut y elevación de los puntos objetivos generados a partir de pasos de azimut y elevación iguales a 8 y 1 respectivamente. Ahora, en cuanto a la intensidad de colores de la gráfica de dispersión, se observa que existe mayor intensidad del color azul (distorsiones muy bajas) en los rangos de azimut donde no contienen a los azimuts de los oídos. Esto porque los rangos corresponden de acuerdo al sistema de referencia de coor-

Figura 3.8: Gráfico de Scatter (azimut vs elevación vs Distorsión Espectral media) para los 4 métodos.

denadas esféricas a la posición frontal de la cabeza humana, donde lógicamente existirá menores distorsiones, ya que las HRTFs medidas no tendrán obstáculos para que los oídos las perciban claramente, mientras que en los rangos de azimut que contienen las ubicaciones de los oídos se encuentran con un color azul mucho más claro (mayor intensidad del celeste), donde es evidente una distorsión espectral más elevada por el efecto sombra de la cabeza humana mencionada anteriormente.

En cuanto a la Figura 3.10, también tiene similitud con las Figuras 3.8 y 3.9, solo que en esta ocasión considera los azimuts y elevaciones de los puntos objetivos derivados de los pasos de azimut y elevación iguales a 8 y 5 respectivamente. En consecuencia, al analizar la escala de colores en los rangos de azimut que no contienen a los oídos, existe color celeste, cuyos valores de distorsión se aproximan a 3dB junto con una pequeña escala de azul intenso que representa valores de distorsión espectral entre 0 y 2dB. Esto porque los rangos de azimut corresponden a la posición frontal de la cabeza humana donde no existirá obstáculos, i.e., el efecto sombra de la cabeza, mientras que en los rangos de azimut que contienen las ubicaciones de los oídos del sujeto, se tiene una escala de colores amarillo y rojo, las cuales poseen distorsiones espectrales que van por encima de los 3dB hasta los 8dB, a causa del efecto sombra de la cabeza.

También cabe recalcar que la Figura 3.10, solo está comparando tres de los cuatro métodos de interpolación, en donde la lineal adyacente no fue tomada en cuenta, debido a que exige un paso de azimut o de elevación igual a 1.

Ya concluido el análisis de la distorsión espectral, considerando las ubicaciones de los puntos objetivos generados por los pasos de azimut y elevación. Se procede al análisis de la

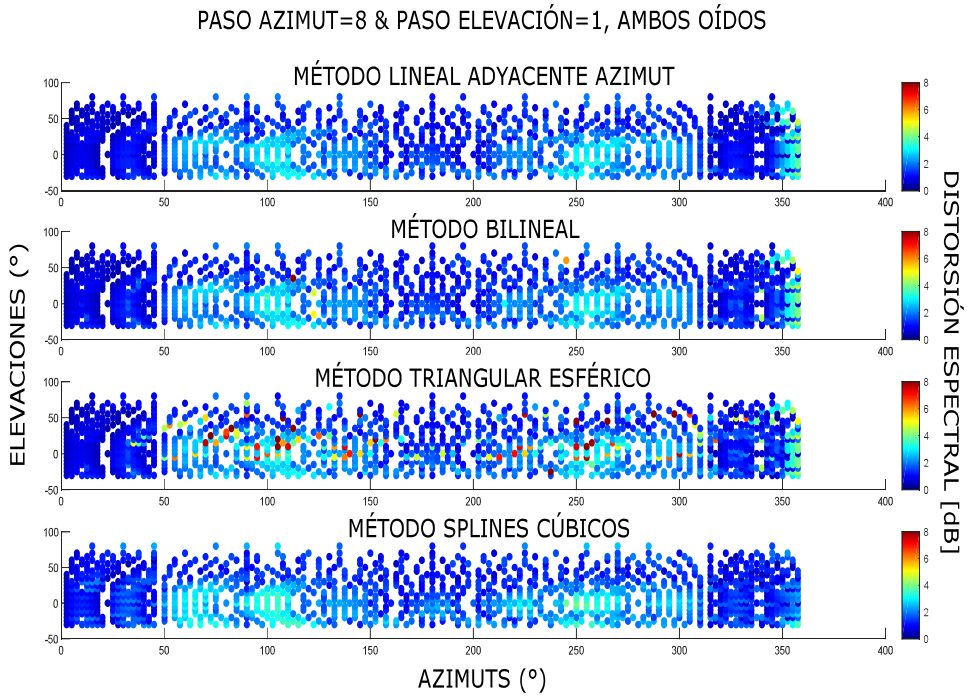

Figura 3.9: Gráfico de Scatter (azimut vs elevación vs Distorsión Espectral media) para los 4 métodos.

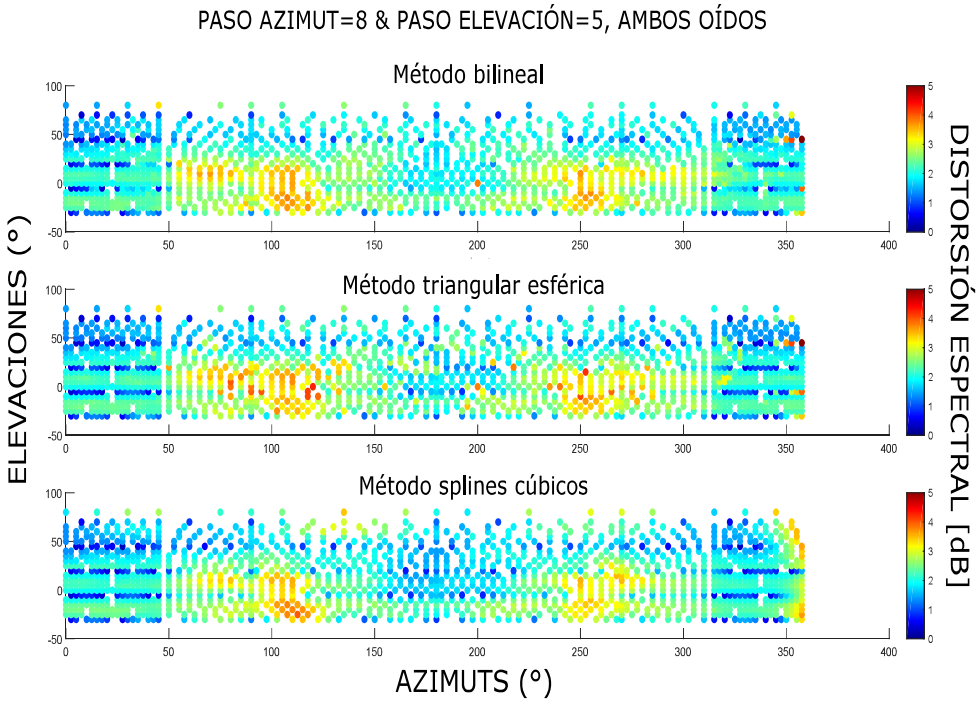

Figura 3.10: Gráfico de Scatter (azimut vs elevación vs Distorsión Espectral media) para los 3 métodos.

misma métrica, pero esta vez tomando en cuenta los planos ipsilateral, contralateral y me-

diano.
Antes que nada, cabe resaltar que para este nuevo análisis, también se utiliza nuevamente las 3 combinaciones de pasos, debido a que no se puede comparar los cuatro métodos en todos los pasos.
Continuando con el análisis, se observa la Figura 3.11.a, la cual indica un gráfico de barras que relaciona cada uno de los cuatro métodos en cada plano con la distorsión espectral, además de indicar sus respectivas variaciones mediante las desviaciones estándar, donde a su vez considera la primera combinación de pasos de azimut y elevación iguales a 1 y 5 respectivamente. Dicho esto, se observa que el plano que presenta mayor distorsión espectral es el contralateral. Esto porque el plano agrupa los puntos objetivos (HRTFs interpoladas) cuyos azimuts se encuentren entre $0° < \theta < 180°$ con respecto al oído derecho y entre $180° < \theta < 360°$ con respecto al oído izquierdo. En otras palabras, agrupa únicamente los puntos objetivos o HRTFs interpoladas, cuyas direcciones de medición de las HRTFs sean contrarias a las ubicaciones de los oídos, de manera que estas distorsiones espectrales son producto del efecto sombra de la cabeza. En cuanto a las distorsiones y desviaciones estándar a nivel de métodos de interpolación, se observa una similitud entre los cuatro métodos. La explicación del resultado obtenido es porque la presente combinación de pasos causa que en la mayoría de métodos de interpolación se resuman a una interpolación lineal adyacente.

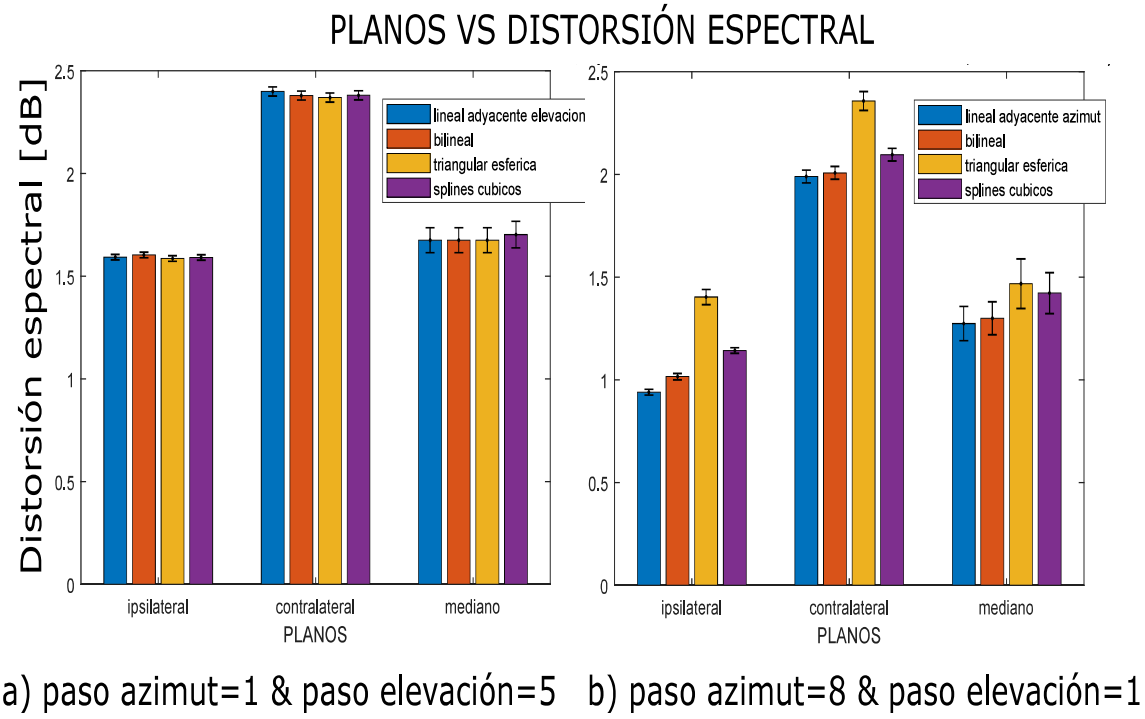

Figura 3.11: Gráfico de barras (PLANOS vs Distorsión espectral) para los 4 métodos. Las barras de error representan ± 2 desviaciones estándar.

Con respecto a la Figura 3.11.b, muestra un resultado similar a la Figura 3.11.a, ya que sus altos niveles de distorsión espectral se encuentran también en el plano contralateral por la misma razón explicada anteriormente. Sin embargo, las distorsiones espectrales en cada

uno de los métodos ya son diferentes entre sí, al igual que las desviaciones estándar que únicamente son las variaciones de las distorsiones con respecto a la distorsión espectral media, de manera que el método con distorsiones espectrales bajas es la lineal adyacente, mientras que el método que posee mayores niveles de distorsión es la triangular esférica.
Por otra parte, la Figura 3.12, al igual que en las Figuras 3.11.a y 3.11.b, presenta un gráfico de barras que relaciona los métodos de interpolación en cada plano con la distorsión espectral, pero considerando pasos de azimut y elevación iguales a 8 y 5 respectivamente. Por tanto se obtiene resultados similares a las anteriores figuras, donde las distorsiones espectrales son elevadas en el plano contralateral, debido al efecto sombra de la cabeza que se mencionó anteriormente y también porque se está utilizando pasos de azimut y elevación grandes que conllevan a una distorsión mucho mayor (caso extremo), i.e., los puntos vecinos se distancian más de los objetivos, haciendo que los errores de aproximación aumenten.

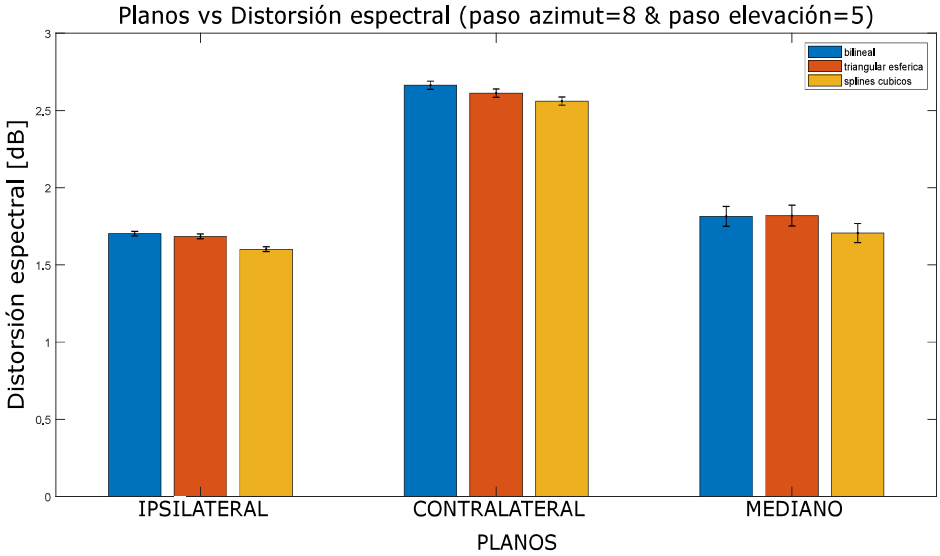

Figura 3.12: Gráfico de barras (PLANOS vs Distorsión espectral) para los 3 métodos. Las barras de error representan ± 2 desviaciones estándar.

Por último, se analiza las distorsiones en base a las cuatro bandas de frecuencia, cuyo rango total va desde los 200Hz hasta los 14KHz. Además de que se utiliza nuevamente las tres combinaciones de pasos por las razones mencionadas anteriormente.
Dicho esto, se procede a analizar la Figura 3.13.a, cuya combinación de pasos de azimut y elevación son 1 y 5 respectivamente, de manera que la banda cuyos niveles de distorsión espectral son bajos, es la de $4Khz < f < 8Khz$, debido a que el suavizado espectral reduce la mayor cantidad de picos o fluctuaciones en las HRTFs por posible filtración del ruido.
En cuanto a la Figura 3.13.b, la cual considera pasos de azimut y elevación iguales a 8 y 1 respectivamente, se observa que los métodos en bajas frecuencias, la distorsión espectral es menor, mientras que en las altas frecuencias, en cambio son altas, de manera que resulta difícil deducir cuales bandas tienen menor y mayor distorsión espectral.
Finalmente, la Figura 3.14 indica los niveles de distorsión espectral en las bandas de fre-

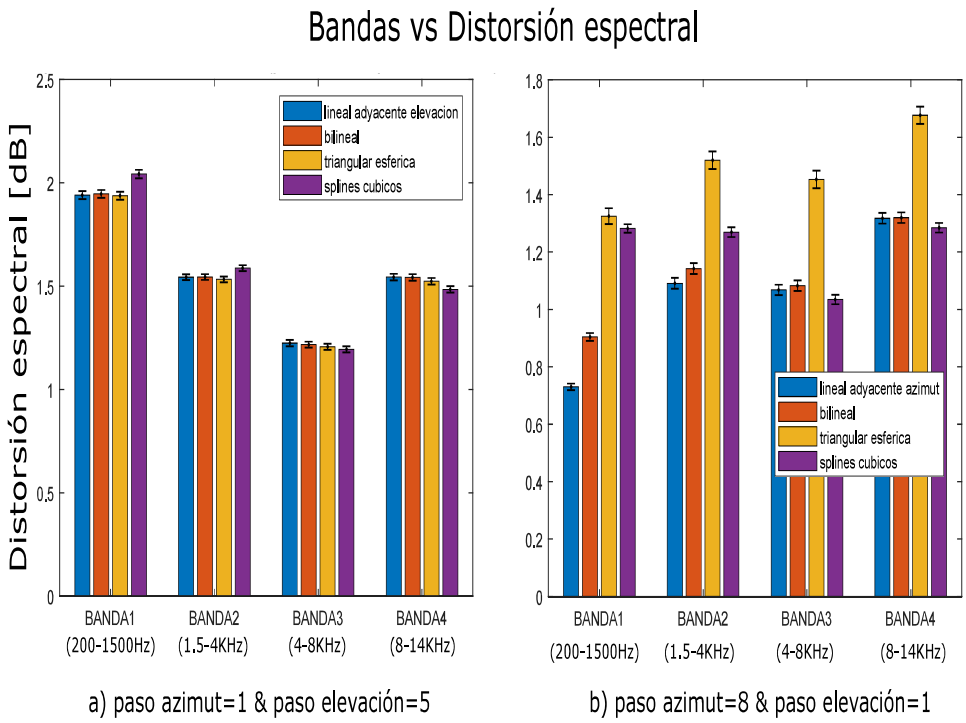

Figura 3.13: Gráfico de barras (BANDAS vs Distorsión espectral) para los cuatro métodos. Las barras de error representan ± 2 desviaciones estándar.

cuencia para la combinación de pasos de azimut y elevación iguales a 8 y 5 respectivamente. Con respecto a la banda que presenta bajos niveles de distorsión espectral, es la de $4KHz < f < 8KHz$, debido a la explicación mencionada anteriormente.

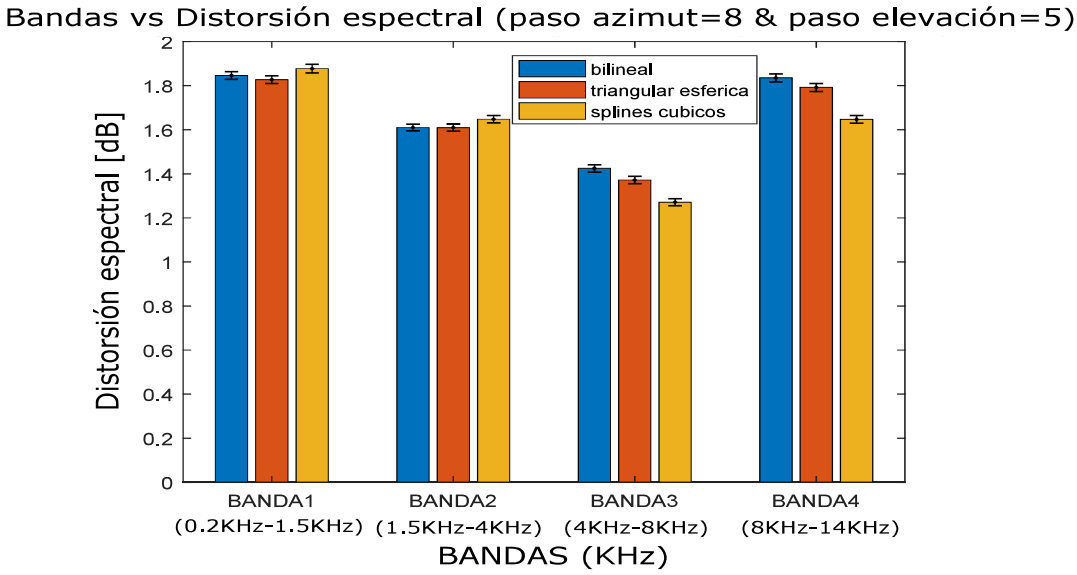

Figura 3.14: Gráfico de barras (BANDAS vs Distorsión espectral) para los 3 métodos. Las barras de error representan ± 2 desviaciones estándar.

4. CONCLUSIONES Y RECOMENDACIONES

4.1. CONCLUSIONES

En el presente estudio técnico se realiza un análisis comparativo de cuatro métodos de interpolación de HRTFs basados en ponderaciones a través de simulaciones en Matlab, para lo cual, dichas comparaciones se realizan mediante la utilización de la métrica de distorsión espectral para diferentes parámetros como suavizado espectral, sujetos, dirección y pasos de azimut y/o elevación (i.e. número de HRTFs interpolantes).

Luego de analizar la utilización del suavizado espectral en los cuatro métodos de interpolación, se concluye que las interpolaciones resultan más eficientes al utilizar el suavizado espectral, debido a que reduce considerablemente los errores de aproximación, los cuales a su vez conducen a menores distorsiones espectrales. En otras palabras, el suavizado espectral tiene como objetivo simplificar las señales para que las funciones interpoladas sean menos difíciles de aproximar a las HRTFs verdaderas.

Es importante recordar que los archivos SOFA con los cuales creamos la base de datos para la aplicación de los cuatro métodos de interpolación, es no uniforme. En consecuencia, se propone diferentes modificaciones a cada uno de los métodos de interpolación lineal adyacente, bilineal y triangular esférica para solucionar el problema presente en las interpolaciones. Entonces, se presenta como solución a la interpolación lineal adyacente, el uso de las distancias euclidianas, debido a que la base de datos es no uniforme en elevación. Por otro lado, la bilineal y triangular esférica, al no poder utilizar las fórmulas para el cálculo de sus respectivos pesos a causa de la no uniformidad presente en las cuadrículas o triángulos de medición, se utiliza el concepto de razones entre áreas en un nuevo plano proyectado utilizando PCA.

En cuanto a la métrica de distorsión espectral, se concluye que a causa del efecto sombra de la cabeza, se producen elevados niveles de dicha métrica en el plano contralateral, mientras que en los planos ipsilateral y mediano, contienen bajos niveles de distorsión espectral, debido a que en dichos planos no existe tal efecto.

Por otra parte, luego de realizar diferentes análisis de la distorsión espectral en parámetros como: los pasos de azimut y elevación, en función de la dirección, con y sin la utilización del suavizado espectral, se concluye que al interpolar en los pasos más bajos (i.e. paso elevación o azimut igual a 1 o 2), el mejor método de interpolación es la lineal adyacente, mientras que en pasos altos (e.g. en el caso extremo de paso azimut=8 & paso elevación=5), y con la utilización de las soluciones propuestas anteriormente para los métodos bilineal, triangular esférica y splines cúbicos, se concluye que las desviaciones estándar en los 3 métodos son similares, i.e., presentan diferencias de distorsión espectral despreciables. Por consiguiente, la metodología propuesta basada en PCA es una alternativa viable de interpolación en mallas no uniformes.

4.2. RECOMENDACIONES

Con el objetivo de seguir investigando el presente tema, se recomienda aplicar estos métodos en diferentes escenarios. Por ejemplo, sería recomendable utilizar estos métodos en una base de datos cuya malla de mediciones de HRTFs sea totalmente uniforme, donde se espera tener niveles de distorsión espectral mucho más bajos que los del presente estudio.

En cuanto a otros posibles trabajos a futuro, podría ser el análisis de los métodos de interpolación para las HRTFs de campo próximo.

Finalmente, otro trabajo sería el de analizar métodos de interpolación de HRTFs basados en reducción de dimensionalidad lineal (e.g. PCA) y no lineal (e.g. autoencoders, Isomap [4])

5. REFERENCIAS BIBLIOGRÁFICAS

[1] B. Xie, R. Dr. N. Xiang, and J. Blauert, *Head-Related Transfer Function and Virtual Auditory Display: Second Edition*, ser. A Title in J. Ross Publishing's Acoustics: Information and Communication Series. S. Pine Island Rd.: J. Ross Publishing, 2013. [Online]. Disponible en: https://books.google.com.ec/books?id=LEaNmQEACAAJ

[2] MediaWiki, "SOFA (spatially oriented format for acoustics)," citado el 26/03/2019. [Online]. Disponible en: https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)

[3] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, no. 3, pp. 267–280, Jun. 2007. [Online]. Disponible en: https://doi.org/10.1007/s11433-007-0018-x

[4] F. Grijalva, L. C. Martini, D. Florencio, and S. Goldenstein, "Interpolation of Head-Related Transfer Functions Using Manifold Learning," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 221–225, Feb. 2017.

[5] A. Kohlrausch, J. Braasch, D. Kolossa, and J. Blauert, "An Introduction to Binaural Processing," in *The Technology of Binaural Listening*, ser. Modern Acoustics and Signal Processing, J. Blauert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–32. [Online]. Disponible en: https://doi.org/10.1007/978-3-642-37762-4_1

[6] J. C Schacher and P. Kocher, "Ambisonics spatialization tools for max/msp," *Omni*, Jan. 2006.

[7] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3, pp. 171–218, Jan. 1992. [Online]. Disponible en: http://www.sciencedirect.com/science/article/pii/0003682X9290046U

[8] J. Lucio Naranjo, "Inteligência computacional aplicada na geração de respostas impulsivas bi-auriculares e em aurilização de salas," Ph.D. dissertation, Universidade do Estado do Rio de Janeiro, Rio de Janeiro,Brasil, May 2014.

[9] Wikipedia, "Oído," Mar. 2019. [Online]. Disponible en: https://es.wikipedia.org/w/index.php?title=O%C3%ADdo&oldid=114412311

[10] H. Fletcher, "Auditory patterns," *Reviews of Modern Physics*, vol. 12, no. 1, pp. 47–65, Jan. 1940. [Online]. Disponible en: https://link.aps.org/doi/10.1103/RevModPhys.12.47

[11] Alamy, "El análisis de las frecuencias de sonido por la membrana basilar," citado el 21/02/2019. [Online]. Disponible en: https://www.alamy.es/foto-el-analisis-de-las-frecuencias-de-sonido-por-la-membrana-basilar-24898700.html

[12] B. C. J. Moore, *An Introduction to the Psychology of Hearing: Sixth Edition*. Boston: BRILL, 2012.

[13] L. Rayleigh, "XII. On our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, Feb. 1907. [Online]. Disponible en: https://doi.org/10.1080/14786440709463595

[14] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, Jun. 2014. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.4875714

[15] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, May 2002. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.1471898

[16] J. C. Middlebrooks and D. M. Green, "Sound Localization by Human Listeners," *Annual Review of Psychology*, vol. 42, no. 1, pp. 135–159, 1991. [Online]. Disponible en: https://doi.org/10.1146/annurev.ps.42.020191.001031

[17] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, Apr. 1999. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.426899

[18] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2325–2332, Oct. 1997. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.419642

[19] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, Feb. 2001. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.1349185

[20] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007.

[21] UCDAVIS, "Spatial sound research," citado 27/03/2019. [Online]. Disponible en: https://www.ece.ucdavis.edu/cipic/spatial-sound/research/

[22] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, Apr. 2002. [Online]. Disponible en: https://orbi.uliege.be/handle/2268/34825

[23] P. Majdak, P. Balazs, and B. Laback, “Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions,” *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, pp. 623–637, Jul. 2007. [Online]. Disponible en: http://www.aes.org/e-lib/browse.cfm?elib=14190

[24] S. Müller and P. Massarani, “Distortion inmunity in impulse response measurements with sweeps,” *Congreso Internacional de Sonido y Vibración*, p. 9, 2011.

[25] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, Oct. 2001, pp. 99–102.

[26] A. R. Institute, “About ARI,” citado 12/04/2019. [Online]. Disponible en: https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=category&id=203&Itemid=643&lang=en

[27] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR,” *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.412407

[28] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, “Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions.” Audio Engineering Society, May 2013. [Online]. Disponible en: http://www.aes.org/e-lib/browse.cfm?elib=16781

[29] K. Vijayan, V. Kumar, and K. S. R. Murty, “Allpass modelling of Fourier phase for speaker verification,” *Speaker Odyssey*, p. 9, 2014.

[30] D. J. Kistler and F. L. Wightman, “A model of head□related transfer functions based on principal components analysis and minimum□phase reconstruction,” *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, Mar. 1992. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.402444

[31] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, “On the minimum-phase approximation of head-related transfer functions,” in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, Oct. 1995, pp. 84–87.

[32] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models: Second Edition*. Springer Science & Business Media, Mar. 2013.

[33] Wikipedia, “Interpolación,” Sep. 2018. [Online]. Disponible en: https://es.wikipedia.org/w/index.php?title=Interpolaci%C3%B3n&oldid=110875691

[34] Ángel Fernando García, “Interpolación con MATLAB. Extrapolación,” 2016. [Online]. Disponible en: http://www.sc.ehu.es/sbweb/fisica3/datos/ajuste/interpol_matlab.html

[35] F. Grijalva, “Manifold learning for spatial audio rendering (aprendizado de variedades para a sintese de audio espacial),” Ph.D. dissertation, University of Campinas, São Paulo, Brasil, 2018.

[36] Wikipedia, “Variedad (matemáticas),” Mar. 2019. [Online]. Disponible en: https://es.wikipedia.org/w/index.php?title=Variedad_(matem%C3%A1ticas)&oldid=114905081

[37] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, Feb. 2010. [Online]. Disponible en: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.101

ANEXOS

ANEXO A. CÓDIGO FUENTE MATLAB

El código se encuentra en el CD adjunto a este documento

ANEXO B. Descripción de los 47 campos de los archivos SOFA

ANEXO B: Descripción de los 47 campos de los archivos SOFA

Nombre	Descripción
HRTF.GLOBAL_Conventions	'SOFA'
HRTF.GLOBAL_Version	'1.0'
HRTF.GLOBAL_SOFAConventions	'SimpleFreeFieldHRIR'
HRTF.GLOBAL_SOFAConventionsVersion	'1.0'
HRTF.GLOBAL_APIName	'ARI
HRTF.GLOBAL_APIVersion	'0.4.3'
HRTF.GLOBAL_ApplicationName	'AMTatARI'
HRTF.GLOBAL_ApplicationVersion	'3.1.10'
HRTF.GLOBAL_AuthorContact	'piotr@majdak.com;michael.mihocic@oeaw.ac.at'
HRTF.GLOBAL_Comment	1x341 char
HRTF.GLOBAL_DataType	'FIR'
HRTF.GLOBAL_History	1x44 char
HRTF.GLOBAL_License	'Creative Commons Attribution-ShareAlike 3.0 Unported License'
HRTF.GLOBAL_Organization	'Acoustics Research Institute, Austrian Academy of Sciences'
HRTF.GLOBAL_References	'Majdak, P., Goupell, M. J., and Laback, B. (2010). "3-D localization of virtual sound sources: effects of visual environment, po...' <Preview truncated at 128 characters>
HRTF.GLOBAL_RoomType	'free field'
HRTF.GLOBAL_Origin	'Acoustically measured'
HRTF.GLOBAL_DateCreated	'2014-04-24 14:16:58'
HRTF.GLOBAL_DateModified	'2014-04-24 14:16:58'
HRTF.GLOBAL_Title	'HRTF'
HRTF.GLOBAL_DatabaseName	'ARI'
HRTF.GLOBAL_ListenerShortName	NH2
HRTF.GLOBAL_ReceiverDescription	'In-the-ear microphones (KE-4-211-2, Sennheiser) connected via amplifiers (FP-MP1, RDL) to the digital audio interface (ADI-8, RM...' <Preview truncated at 128 characters>
HRTF.GLOBAL_RoomDescription	'Customized IAC semi-anechoic room (6.2m 5.5m 2.9m)'
HRTF.GLOBAL_RoomLocation	'Vienna, Austria'
HRTF.GLOBAL_SourceDescription	'Twenty-two loudspeakers (the variation in the frequency response was +/- 4 dB in the range from 200 to 16000 Hz) were mounted at...' <Preview truncated at 128 characters>
HRTF.GLOBAL_EmitterDescription	'VIFA 10 BGS'
HRTF.API	1x1 struct
HRTF.ListenerPosition	[0,0,0]
HRTF.ListenerPosition_Type	'cartesian'
HRTF.ListenerPosition_Units	'metre'
HRTF.ReceiverPosition	[0,-0.090000000000000,0;0,0.090000000000000,0]
HRTF.ReceiverPosition_Type	'cartesian'
HRTF.ReceiverPosition_Units	'metre'
HRTF.SourcePosition	1550x3 double
HRTF.SourcePosition_Type	'spherical'
HRTF.SourcePosition_Units	'degree, degree, metre'
HRTF.EmitterPosition	[0,0,0]
HRTF.EmitterPosition_Type	'cartesian'
HRTF.EmitterPosition_Units	'metre'
HRTF.ListenerUp	[0,0,1]
HRTF.ListenerView	[1,0,0]
HRTF.ListenerView_Type	'cartesian'
HRTF.ListenerView_Units	'metre'
HRTF.MeasurementSourceAudioChannel	1550x1 double
HRTF.MeasurementAudioLatency	1550x2 double
HRTF.Data	1x1 struct

ORDEN DE EMPASTADO