ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

ANÁLISIS Y SÍNTESIS DE AUDIO ESPACIAL USANDO ANÁLISIS DE COMPONENTES PRINCIPALES

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

BORIS JAVIER ESCOBAR CASTRO

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 Boris Javier Escobar Castro, 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, BORIS JAVIER ESCOBAR CASTRO, 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.

BORIS JAVIER ESCOBAR CASTRO

DEDICATORIA

Este trabajo esta dedicada a:

A mis padres, Víctor Javier y Narcisa de Jesús
quienes con su amor, paciencia y esfuerzo
han sabido inculcar en mí, la valentía y el coraje
necesarios para siempre seguir adelante.

A mis hermanas, Karla y Odalys
por su infinito amor, cariño y confianza
y por llenarme de alegría día tras día.

Finalmente dedico este trabajo de titulación
a todas las personas que estuvieron junto a mí,
aconsejándome, apoyándome y guiándome
durante todo este trayecto.

Boris Javier

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a mis padres y hermanas por su apoyo incondicional, por sus palabras de ánimo y aliento en los momentos que más los he necesitado.

Al PhD. Felipe Grijalva por su diligente trabajo durante todo este proyecto, quien con su ayuda y conocimiento ha sabido guiar de la mejor manera el desarrollo del mismo.

A la Escuela Politécnica Nacional por brindarme el conocimiento necesario requerido para mi desarrollo profesional, a sus docentes por las experiencias transmitidas y conocimientos compartidos.

Finalmente, agradezco a todas las personas que compartieron junto a mi dentro de las aulas de la Escuela Politécnica Nacional

Boris Javier

ÍNDICE DE CONTENIDO

AVAL I
DECLARACIÓN DE AUTORIA II
DEDICATORIA III
AGRADECIMIENTO IV
ÍNDICE DE CONTENIDO V
RESUMEN VII
ABSTRACT VIII
1. INTRODUCCIÓN 1
1.1. OBJETIVOS 2
1.2. ALCANCE 2
1.3. MARCO TEÓRICO 3
1.3.1. AUDIO ESPACIAL 3
1.3.1.1. Sistemas de Coordenadas 4
1.3.1.2. El sistema auditivo 5
1.3.1.3. Filtros Auditivos 6
1.3.1.4. Pistas de localización auditiva 7
1.3.2. HEAD-RELATED TRANSFER FUNCTIONS 9
1.3.2.1. Definición 9
1.3.2.2. Medición de las HRTFs 10
1.3.2.3. Principales características de las HRTFs 13
1.3.2.4. Bases de datos de las HRTFs 15
1.3.2.5. Aproximación de fase mínima de las HRTFs 17
1.3.3. ANÁLISIS DE COMPONENTES PRINCIPALES 18
1.3.3.1. Formulación del problema de PCA 19
1.3.3.2. Algoritmo de PCA 22
1.3.3.3. Reconstrucción a partir de los datos de baja dimensionalidad 24
1.3.3.4. Selección del número de componentes principales 25
1.3.3.5. Análisis de componentes principales para HRTFs 27

2. METODOLOGÍA 29
2.1. BASE DE DATOS ARI 29
2.1.1. EXTRACCIÓN DE DATOS DE LA BASE DE DATOS ARI 30
2.1.1.1. Construcción de tabla de datos para los 95 sujetos estudiados en el proyecto 33
2.2. PRE-PROCESAMIENTO DE DATOS 34
2.2.1. APLICACIÓN DE LA TRANSFORMADA RÁPIDA DE FOURIER 34
2.2.2. APLICACIÓN DE FILTRADO Y SUAVIZADO ESPECTRAL 35
2.3. ANÁLISIS DE HRTFs USANDO PCA 38
2.3.1. USO DE PCA EN MATLAB 38
2.3.2. SELECCIÓN DE NÚMERO DE COMPONENTES PRINCIPALES 39
2.4. SÍNTESIS DE HRTFs CON PCA 41
2.4.1. RECONSTRUCCIÓN DE LAS HRTFs ORIGINALES CON BASE EN LAS HRTFs DE BAJA DIMENSIÓN. 41
2.4.2. K-FOLD CROSS-VALIDATION 43
3. RESULTADOS Y DISCUSIÓN (APLICACIÓN METODOLÓGICA) 46
3.1. RESULTADOS OBTENIDOS EN LA ETAPA DE ANÁLISIS 46
3.2. RESULTADOS OBTENIDOS EN LA ETAPA DE SÍNTESIS 51
4. CONCLUSIONES Y RECOMENDACIONES 57
4.1. CONCLUSIONES 57
4.2. RECOMENDACIONES 58
5. REFERENCIAS BIBLIOGRÁFICAS 59
ANEXOS 61

RESUMEN

El audio espacial es un conjunto de técnicas que tiene como fin simular fuentes de sonido ubicadas en una posición específica del espacio. Esto se consigue gracias a las Funciones de Transferencia Derivadas de la Cabeza (*Head-Related Transfer Functions*,HRTF) que modelan las características anatómicas de una persona y su interacción con el campo sonoro incidente.

El presente proyecto muestra un estudio sobre el análisis y síntesis de HRTFs mediante el análisis de componentes principales (*Principal Component Analysis*, PCA).
Con este objetivo en mente,partimos de una investigación bibliográfica sobre el audio espacial, las HRTFs y PCA. Específicamente se estudia, el sistema auditivo, filtros auditivos, consideraciones prácticas al momento de realizar mediciones de HRTFs, características en tiempo y frecuencia de las HRTFS y cómo aplicar PCA en las mismas.

Enseguida, se explican a detalle los procedimientos de análisis y síntesis efectuados durante este proyecto. En el análisis, se determinó el número de componentes principales que permitan conservar el 95 % de la información de las HRTFs originales. Seguido de esto, se evaluó el desempeño en la síntesis al reconstruir las HRTFs a partir de su representación en baja dimensionalidad.

Los resultados obtenidos muestran que los ángulos de azimut están fuertemente correlacionados con las componentes principales de PCA, mientras que los ángulos de elevación son pobremente descritos por las mismas. Por último, el error de validación cruzada usado en la síntesis muestra que existe mayor distorsión espectral en el ipsilateral del oyente.

Finalmente, y considerando la distorsión espectral obtenida en la síntesis concluimos que PCA captura de forma correcta información relevante de las HRTFs.

PALABRAS CLAVE: audio espacial, HRTF, PCA, distorsión espectral.

ABSTRACT

Spatial audio is a set of techniques that aims to simulate sound sources located in a specific position of space. This is achieved thanks to the HRTFs that model the anatomical characteristics of a person and their interaction with the incident sound field.

This project presents a study on the analysis and synthesis of Head-Related Transfer Functions (HRTFs) through Principal Component Analysis (PCA).

With this aim, we start with a bibliographical research on spatial audio, HRTFs and PCA. Specifically, we study the auditory system, auditory filters, practical considerations when performing HRTFs measurements, time and frequency characteristics of HRTFS and how to apply PCA on them.

Next, we explain in detail the analysis and synthesis procedures carried out during this project . In the analysis stage, we determine the number of principal components that preserves 95 % of variance of the high-dimensional HRTFs. Following this, we evaluate the performance in a synthesis stage by reconstructing the HRTFs from their low-dimensional representation.

The results obtained show that the azimuth angles are strongly correlated with the principal components of PCA, while the elevation angles are poorly described by them. Lastly, the cross-validation error used in the synthesis shows that there is greater spectral distortion in the listener's ipsilateral.

Finally, and considering the spectral distortion obtained in the synthesis, we conclude that PCA correctly captures relevant information from the HRTFs.

KEYWORDS: spatial audio, HRTF, PCA, Spectral distortion.

1. INTRODUCCIÓN

El sonido es uno de los primeros estímulos que experimenta el ser humano al nacer. Pasar del tranquilo entorno del vientre de una madre, a un ambiente altamente caótico en lo que a sonidos se refiere, es un cambio bastante drástico. Es por eso que el oído humano aprende a discriminar sonidos desde una temprana edad, haciendo que este ejercicio sea particular para cada individuo, ya que las características fisiológicas de cada persona son únicas, e influyen de forma directa en como percibimos el sonido. Hoy en día es común encontrar sonidos en dispositivos portátiles y auriculares, en los cuales el audio binaural se puede presentar de forma conveniente. El audio binaural espacial (o audio espacial para abreviar) tiene como objetivo crear una experiencia de audio en tres dimensiones, en la cual el oyente pueda localizar una fuente sonora que se encuentra en una ubicación específica. Esto lo consigue con base en las Funciones de Transferencia Derivada de la Cabeza (*Head-Related Transfer Function*, HRTF) [1].

Las HRTFs son funciones de transferencia medidas que describen los patrones de directividad de los oídos humanos, es decir, una descripción de cómo el sonido, al llegar de una dirección dada, interactúa con el oído izquierdo y el oído derecho. Dado que las HRTF se modelan en base a las alteraciones y característica propias de cada persona (e.g. cabeza, torso, pabellón del oído, etc), estas deberían ser modeladas específicamente para cada individuo, con el fin de garantizar calidad en el audio espacial [1]. Sin embargo, esto resulta ser impráctico, ya que requeriría de un equipo especializado (propiamente se requeriría una cámara anecoica) para la obtención de HRTFs personalizadas [2]. Sumando a esto es importante destacar que la carga computacional y la memoria que requiere la base de datos de una HRTF presentan ser un desafío para la aplicación práctica de audio espacial. Es por este motivo que se requiere una representación más compacta de las HRTFs. Es dentro de este contexto que el presente proyecto usará el Análisis de Componentes Principales ("Principal Component Analysis, PCA"), como herramienta para reducir la complejidad de las HRTFs.

El análisis de componentes principales tiene como objetivo principal transformar un conjunto de variable originales (conocidas también como variables de alta dimensión), en un nuevo conjunto de variables denominadas componentes principales(conocidas también como variables de baja dimensión). Estas últimas se caracterizan por ser variables no correlacionadas entre sí, y además pueden ser ordenadas con base en la cantidad de información que lleva cada componente. Como medida de la cantidad de información que lleva cada componente se usa su varianza. Así, entre mayor sea la varianza mayor será la cantidad de información que tenga dicha componente. Dado que los componentes se ordenan por su cantidad de varianza, esta resulta ser una técnica útil para reducirla dimensionalidad de un conjunto de datos

1.1. OBJETIVOS

El objetivo general de este Proyecto Técnico es Analizar y Sintetizar las Funciones de Transferencia Derivadas de la Cabeza usando Análisis de Componentes Principales.

De este objetivo general, se desprenden los siguientes objetivos específicos:

- Estudiar los conceptos de audio espacial y HRTFs.
- Estudiar los conceptos de análisis de componentes principales.
- Implementar Análisis de Componente Principales en MATLAB para reducir la dimensionalidad de las HRTFs.
- Analizar las HRTFs de baja dimensión, para determinar cómo las componentes principales se relacionan con los ángulos de azimut y elevación.
- Usar PCA para la síntesis de las funciones de transferencia a partir de las HRTFs de baja dimensionalidad

1.2. ALCANCE

Usando las herramientas de MATLAB se llevarán a cabo el análisis y síntesis de las Funciones de Transferencia Derivadas de la Cabeza ("Head-Related Transfer Function, HRTF") para lo cual se empleará el Analisis de componentes principales (*Principal Analysis Component*,PCA) [3]. Para esto se procederá a realizar básicamente tres etapas, en las que se usarán como base las HRTFs en formato SOFA [4] ("Spatially Oriented Format for Acoustics") [4].

- **Preprocesamiento de las HRTFs:** Previo al análisis y manipulación de las HRTFs con PCA es necesario que las HRTFs originales pasen por un banco de filtros auditivos, ya que la resolución de frecuencia del sistema auditivo está relacionada con la forma y el ancho de los mismos [1]. Además, se usará aproximación de fase mínima [1] para las HRTFs.
- **Reducción de Dimensionalidad (Análisis):** En esta etapa se aplicará PCA a fin de reducir la dimensionalidad de las HRTFs originales. Además, se usará como indicador, el porcentaje de varianza retenida para determinar el número de componentes principales que se conservaran luego de la reducción de dimensionalidad. Dentro de esta etapa se analizará las HRTFs de baja de baja dimensión (aquellas obtenidas luego de aplicar PCA) para determinar su relación con los ángulos de elevación y azimut.
- **Sintesis:** En la síntesis se buscará reconstruir las HRTFs de alta dimensión a partir de las HRTFs de baja dimensión (i.e. las obtenidas con PCA).

Por otro lado, se contará con un conjunto de datos de prueba, que estará conformado por un grupo de funciones HRTFs a las cuales no se las procesará con PCA. Esto a fin de que, este conjunto de prueba sea considerado como el valor verdadero para las funciones HRTFs, para en base a este, calcular las métricas de distorsión espectral, y compararlas con las obtenidas luego de la etapa de síntesis.

1.3. MARCO TEÓRICO

En esta sección se abordarán esencialmente tres grandes temas. En la Sección 1.3.1 centraremos nuestra atención en el estudio de audio espacial, dentro del cual revisaremos conceptos, definiciones y principios esenciales de la audición espacial. Por otra parte, en la Sección 1.3.2 se analizarán las Funciones de transferencia relacionadas con la cabeza (*Head-Related Transfer Function*, HRTF). Finalmente, en la Sección 1.3.3 dirigiremos nuestro estudio hacia el Análisis de Componentes Principales (PCA), partiendo de la formulación del problema y el algoritmo usado por PCA, por ultimo, dentro de esta sección se analizarán, el algoritmo para el análisis de componentes principales asociada a las HRTFs.

1.3.1. AUDIO ESPACIAL

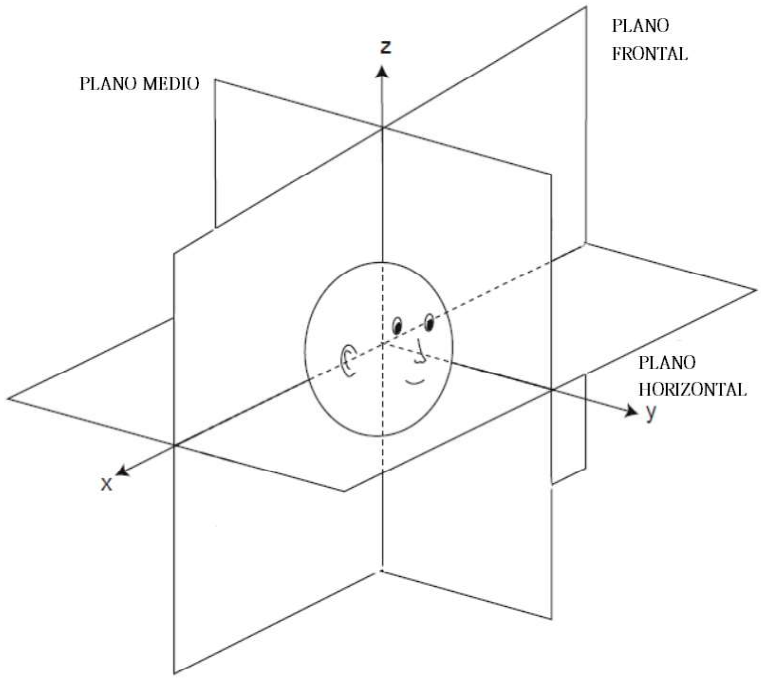

Figura 1.1: Planos de referencia. Adaptado de [1].

Al hablar de audio espacial nos referimos al conjunto de técnicas que modelan la influencia que tienen las características anatómicas (la cabeza, el torso, etc) en el campo sonoro incidente en los oídos [1]. El audio espacial tiene como objetivo crear una experiencia de

audio en tres dimensiones (i.e. audio 3D [5]), en el cual el oyente pueda localizar una fuente sonora que se encuentra en una ubicación específica en el espacio.
Para este fin, existen diferentes formas de generar audio espacial (e.g *Panning* [6], Binaural, y *Sound Field Synthesis* [7], HRTF [2]). Para el presente proyecto se usará la técnica de síntesis de audio espacial basado en HRTF [2].

1.3.1.1. Sistemas de Coordenadas

Para un oyente que se encuentra en un entorno en tres dimensiones, la posición de una fuente de sonido puede ser especificada en términos de la dirección y la distancia con respecto a la cabeza del oyente. Usualmente, en el estudio del audio espacial el centro de la cabeza del oyente es usado como el origen del sistema de coordenadas [1].

Así, para definir la localización de una fuente sonora, es importante señal los tres planos de referencia (ver Figura 1.1), que se usarán a lo largo de este proyecto:

- Plano medio: plano y-z
- Plano frontal: plano x-z
- Plano horizontal o transversal: plano x-y

Se pueden adoptar varios sistemas de coordenadas al describir la posición de una fuente de sonido, sin embargo, dentro del estudio de audio espacial se utilizan comúnmente los sistemas de coordenadas esféricas y coordenadas polares interaurales.

Coordenadas Esféricas

Las coordenadas esféricas especifican una ubicación en base a tres coordenadas (r, θ, ϕ). El radio r representa la distancia a la que se encuentra la fuente sonora medido desde el origen de coordenadas, siendo siempre un número positivo $0 \leq r \leq +\infty$. El ángulo ϕ representa el ángulo de elevación, que varía entre $-90° \leq \phi \leq 90°$, en el que $-90°$ y $90°$ representan las posiciones inferior y superior del oyente respectivamente, mientras que $0°$ representa posiciones en el plano horizontal. El ángulo θ, el cual varía entre $0° \leq \theta \leq 360°$, representa el ángulo de azimut, donde $0°, 90°, 180°$ y $270°$, representan las direcciones frontal, izquierda, trasera y derecha en el plano horizontal, respectivamente. La Figura 1.2 muestra el sistema de coordenadas esféricas.

Coordenadas polares interaurales

El sistema de coordenadas polares interaurales mostrado en la Figura 1.3 especifica una ubicación en base a tres coordenadas (r, θ, ϕ). La definición y rango para r es la misma que en el sistema de coordenadas esféricas (i.e. $0 \leq r \leq +\infty$). El ángulo de azimut θ varía entre $-90° \leq \theta \leq 90°$, y el ángulo de elevación ϕ, varía entre $-90° \leq \phi \leq 270°$. En este sistema

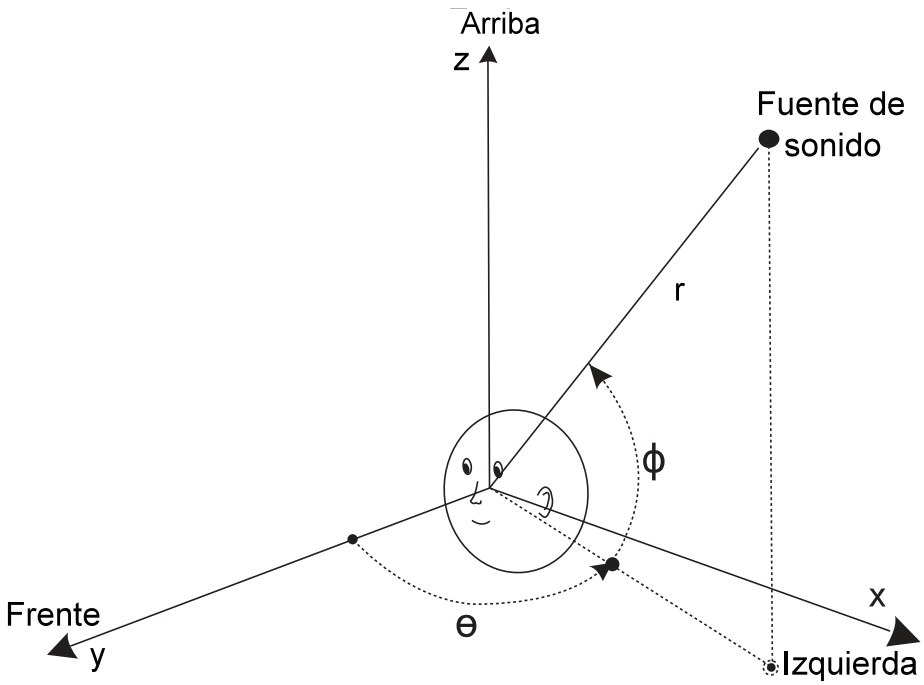

Figura 1.2: Sistema de Coordenadas Esféricas. Adaptado de [1].

las coordenadas

$$(\theta;\phi) = (0°;0°);(0°;90°);(0°;180°);(0°;270°);(90°;0°);(-90°;0°),$$

representan a las posiciones frontal, arriba, atrás, abajo, derecha e izquierda respectivamente.

1.3.1.2. El sistema auditivo

Podemos dividir al sistema auditivo humano en tres partes: parte interna, media y externa del oído (ver Figura 1.4) [1]. El oído externo está compuesto por la pabellón auditivo y el canal auditivo. La forma y dimensiones de la pabellón auditivo varían dependiendo de cada individuo. Por otra parte, el canal auditivo se encarga de transportar y amplificar el sonido hacia el oído medio.

La parte media u oído medio está constituido por el tímpano, la cavidad timpánica, los huesecillos, los músculos y ligamentos asociados, así como la trompa de Eustaquio. El tímpano marca el inicio del oído medio y se encuentra al final del canal auditivo [1]. Los huesecillos dentro de la cavidad timpánica consisten en tres huesos pequeños: el martillo, el yunque y el estribo. Estos a su vez convierten las vibraciones de sonido producidas cuando las ondas de sonido golpean el oído en vibraciones de sonido de mayor presión [1]. El martillo se une al tímpano, el estribo termina en la ventana oval del oído interno y el yunque se encuentra en el medio. La estructura principal en el oído interno se llama cóclea, que convierte las ondas sonoras que llegan al oído interno en impulsos eléctricos, estos son enviados por medio del nervio auditivo al cerebro, quien finalmente traduce estos impulsos en sonido.

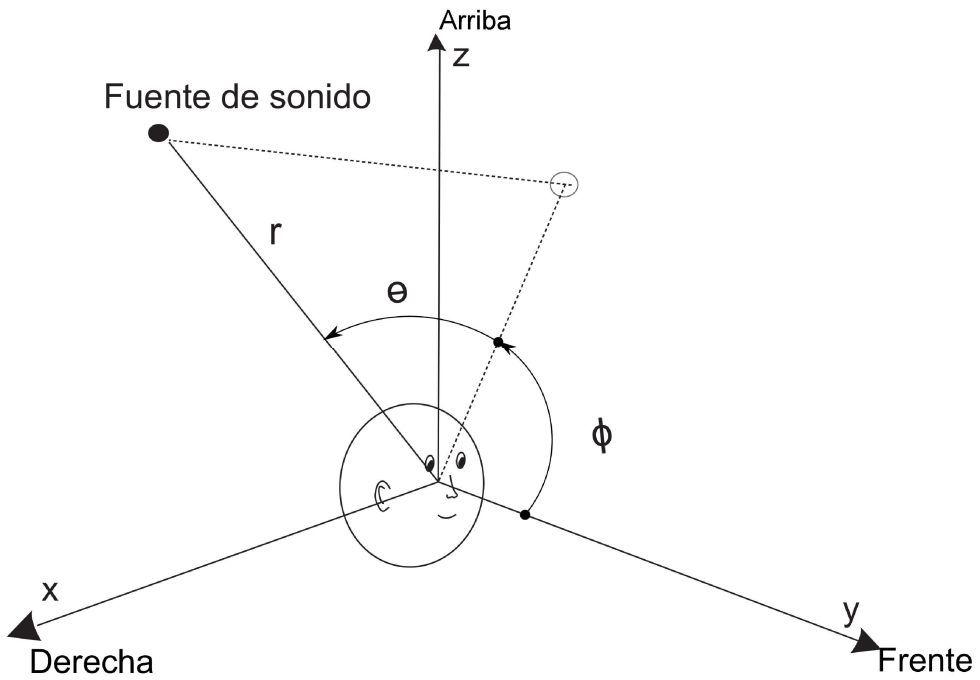

Figura 1.3: Sistema de Coordenadas Polares Interaurales. Adaptado de [1].

1.3.1.3. Filtros Auditivos

La respuesta en frecuencia del oído depende de la membrana basilar [9], la misma que se encuentra dentro de la cóclea y varía en masa y rigidez a lo largo de su longitud, lo que implica que las diferentes frecuencias resuenen particularmente fuerte en diferentes puntos sobre la membrana. Es por esta razón que su comportamiento puede ser descrito como un conjunto de filtros pasa-banda superpuestos con una frecuencia central característica.
Estos filtros se ven caracterizados por el ancho de banda rectangular equivalente (*Equivalent Rectangular Bandwidth* , ERB) [1], que relacionan el filtro auditivo, la frecuencia central y el ancho de banda crítico del filtro. Distintos experimentos psicoacústicos como el de Moore [10] indican que el ERB en Hz está relacionado a una frecuencia central en kHz mediante:

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

Con base en el concepto de filtro auditivo rectangular equivalente, se puede introducir el número de ancho de banda rectangular equivalente (*Equivalent Rectangular Bandwidth Number*, ERBN) [1], que establece el ancho de banda en KHz como

$$\text{ERBN} = 24{,}1 \cdot \log_{10}(4{,}37 \cdot f + 1) \tag{1.2}$$

Es importante destacar que, el ancho de banda del modelo ERB aumenta con la frecuencia, i.e, la resolución de frecuencia del sistema auditivo disminuye a medida que aumenta

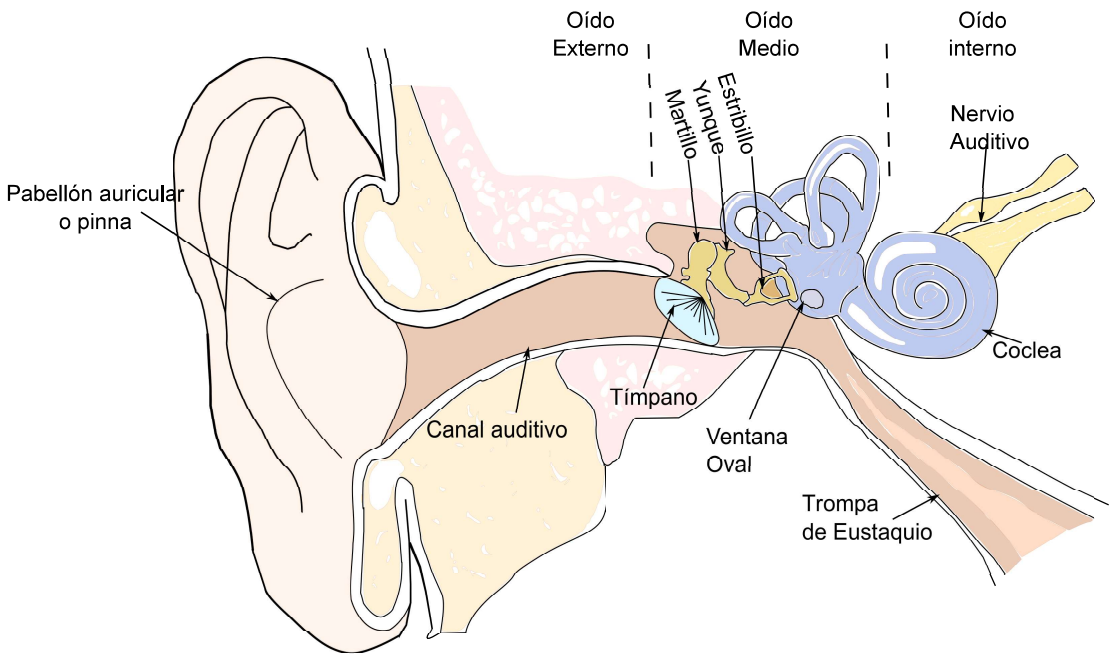

Figura 1.4: Sección transversal del sistema auditivo periférico de los humanos. Adaptado de [8].

la frecuencia.

1.3.1.4. Pistas de localización auditiva

La localización auditiva busca determinar la posición espacial percibida de una fuente sonora en términos de su dirección y distancia en relación con el oyente. Estudios psicoacústicos han demostrado que las pistas de localización direccionales para una única fuente de sonido incluyen la diferencia de tiempo interaural (*Interaural Time Difference*, ITD), la diferencia de nivel interaural (*Interaural Level Difference*, ILD), así como las pistas dinámicas y espectrales [1] .

Diferencia de tiempo interaural

La ITD se refiere a la diferencia de tiempo de llegada de las ondas de sonido entre el oído izquierdo y derecho [1]. Es por ello que, en el plano medio, la ITD es aproximadamente cero, esto debido a que las longitud que recorre el sonido desde la fuente sonora hacia ambos oídos son similares. Así, si una fuente se ubica en el plano medio del oyente, no existirá diferencia en el tiempo de llegada. Sin embargo, si la fuente se ubica en otra posición la ITD será ampliamente más apreciable.

La ITD además es la señal de localización más importante para frecuencias por debajo de aproximadamente 700Hz, [1] dado que los sonidos en este rango de frecuencia presentan longitudes de onda mayores que las dimensiones de una cabeza humana promedio. Esto

nos permite confiar en la información de tiempo (o fase) entre los oídos para determinar la posición de una fuente sonora sin confusión.

Diferencia de nivel interaural

La ILD se define como la diferencia en el nivel de presión sonora entre ambos oídos [1]. Esta toma particular importancia cuando la fuente sonora se desvía del plano medio del oyente, ya que al hacerlo la presión del sonido en el oído más alejado a la fuente sonora (i.e. contralateral) se atenúa, principalmente en frecuencias superiores a los 1500 Hz aproximadamente [1]. Esto ocurre debido a que estas frecuencias presentan longitudes de onda menores que las dimensiones de una cabeza humana promedio. Además, debido al efecto de sombra (i.e. difracción) que produce la cabeza, el sonido en el oído más alejado se atenúa aún más.

Es importante resaltar que, en la zona de transición, entre aproximadamente 700 Hz y 1500 Hz nuestro sistema auditivo usa tanto las diferencias de nivel (ILD) como las diferencias de tiempo (ITD) para la localización en igual medida.

Pistas dinámicas

Hasta el momento contamos únicamente con dos pistas para conocer la ubicación de una fuente sonora. Sin embargo, en conjunto la ITD e ILD resultan ser insuficientes para determinar la posición única de una fuente de sonido.

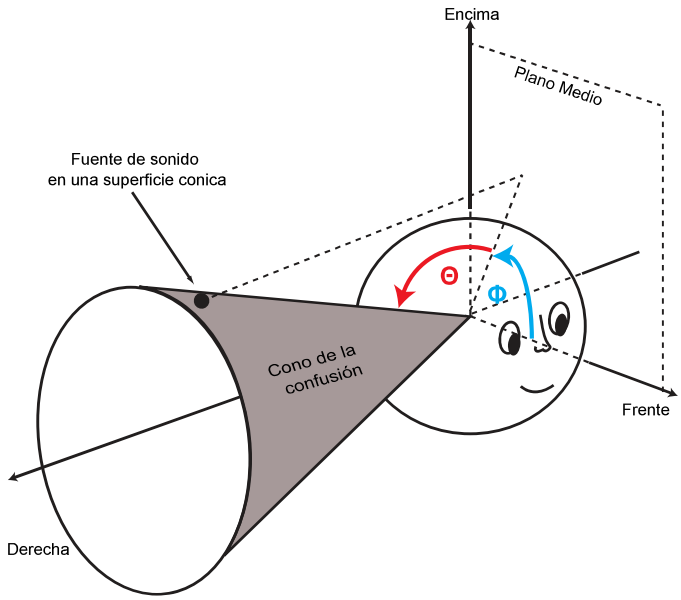

Figura 1.5: Cono de la confusión. Adaptado de [1]

Esto debido a que, en teoría, pueden existir valores iguales de ITD e ILD, para diferentes ubicaciones de una fuente de audio, dentro de una superficie cónica. Considere por ejemplo, si usamos el sistema de coordenadas polares interaurales manteniendo constante el ángulo

de azimut, se forma un cono en cuya superficie los valores de ILD e ITD son teóricamente idénticos. Si bien resulta poco probable que los valores de ITD e ILD sean idénticos en un entorno real con dos fuentes de audio, cuando estas están muy cerca en dos ubicaciones diferentes, pueden surgir ambigüedades.

La Figura 1.5 anteriormente descrita se conoce como: cono de la confusión [1], ya que en la misma se producen confusiones frente-atrás y arriba-abajo . Las confusiones frente-atrás se refieren a la percepción de que un sonido proyectado en una posición adelante del sujeto está detrás del mismo, y viceversa. El mismo concepto se aplica a las confusiones arriba-abajo que corresponden a pistas por debajo o sobre la cabeza del oyente. Es por esto que se vuelve indispensable el uso de pistas dinámicas como movimientos de la cabeza, para poder identificar la ubicación real de una fuente de sonido.
Varios estudios han demostrado la efectividad de los movimientos de la cabeza para resolver confusiones frente-atrás y arriba-abajo [11, 12].

Pistas espectrales

La anatomía del cuerpo tiene gran influencia en como percibimos el sonido, es por ello que las características anatómicas como el pabellón auricular, la cabeza y el torso son responsables de las pistas espectrales también conocidas como pistas monoaurales. Estas pistas proporcionan información útil para la localización vertical y la resolución de confusiones frente-atrás y arriba-abajo que surgen sobre posiciones en el cono de confusión.

1.3.2. HEAD-RELATED TRANSFER FUNCTIONS

Las Funciones de Transferencia Derivadas de la Cabeza (*Head-Related Transfer Functions*, HRTF) son funciones de transferencia que describen los patrones de directividad de los oídos, es decir, presenta una descripción de cómo el sonido, al llegar de una dirección dada, interactúa con el oído izquierdo y el oído derecho [1]. Las HRTFs dependen tanto de las pistas binaurales (i.e. ITD e ILD) como de las monoaurales, y dado que estas son características únicas de la anatomía de cada persona, se puede concluir que las HRTFs son únicas para cada individuo [1].
A continuación, se procederá a describir las principales características de las HRTFs.

1.3.2.1. Definición

El sonido emitido por una fuente llega a los oídos luego de interactuar con las características anatómicas del individuo, haciendo que la señal originalmente emitida, se haya refractado y reflejado tanto en la cabeza, el torso y el pabellón auditivo del sujeto. El resultado captado por los oídos es un conjunto de las pistas estudiadas en la Sección 1.3.1.4 (ITD, ILD y pistas espectrales) [1]. Este conjuntos de pistas son modelados a través de las HRTFs. Un par de HRTFs, para el oído izquierdo y derecho, H_L y H_R respectivamente, se definen como

$$H_L(r,\theta,\phi,f,a) = \frac{P_L(r,\theta,\phi,f,a)}{P_0(r,f)}$$
$$H_R(r,\theta,\phi,f,a) = \frac{P_R(r,\theta,\phi,f,a)}{P_0(r,f)}, \tag{1.3}$$

donde, P_L y P_R representan las presiones de sonido en el dominio de frecuencia en el oído izquierdo y derecho, respectivamente, y P_0 representan la presión de sonido de campo libre en el dominio de frecuencia en el centro de la cabeza con la cabeza ausente [1].

Cada HRTF es una función que dependen tanto de la frecuencia como de la posición de la fuente de sonido, esto en términos de distancia r así como, de los ángulos de azimut θ y elevación ϕ [1]. Además, dado que la HRTF es dependiente de las estructuras y dimensiones anatómicas únicas de cada ser humano, estas características anatómicas forman parte de las Ecuaciones 1.3 presentadas anteriormente y son denotadas por la letra *a*.

En sentido estricto, *a* se refiere a un conjunto de parámetros que especifican las dimensiones de las estructuras anatómicas relevantes de cada sujeto.A pesar de que las HRTFs son funciones multivariable, determinadas variables ganan o pierden importancia dependiendo de la ubicación de la fuente de sonido [13]. Así, por ejemplo, si la distancia entre el oyente y la fuente sonora es aproximadamente mayor a 1 metro ($r > 1$ m), se considera que las HRTFs son independientes de la distancia y se las denominan HRTFs de campo lejano. Existen también las HRTFs de campo cercano que son aquellas que dependen de la distancia [1]. Durante este proyecto trabajaremos únicamente con HRTFs de campo lejano, por lo cual, cuando hablemos de HRTFs nos estaremos refiriendo exclusivamente a las HRTF de campo lejano.

Finalmente, es importante definir las HRIRs (Head-Related Impulse Response), que son la contraparte en el dominio del tiempo de las HRTFs. Las HRTFs y las HRIRs están relacionadas por la transformada de Fourier. Generalmente las HRIRs son medidas directamente como respuestas impulsivas.

1.3.2.2. Medición de las HRTFs

Al momento de realizar mediciones de las HRTFs la configuración más típica utilizada para este fin es el usar una sala anecoica o semi-anecoica, con el sujeto colocado en el centro de la sala y alrededor de él un sistema con uno o más altavoces que se pueden mover para medir las HRTFs en varias direcciones [13]. Ahora bien, fijándonos en la Ecuación 1.3 observamos que las funciones HRTFs son continuas. Sin embargo, en la práctica, las HRTF se miden solo para posiciones discretas en el espacio. Las HRTFs para las posiciones no medidas se pueden calcular por interpolación espacial [14, 15].

Las HRIRs de un individuo se obtienen al reproducir una señal analítica (es decir, una señal sintética conocida), la cual es amplificada, para posteriormente ser reproducida dentro de una sala anecoica o semi-anecoica. En dicha sala se mide directamente las HRIRs sin

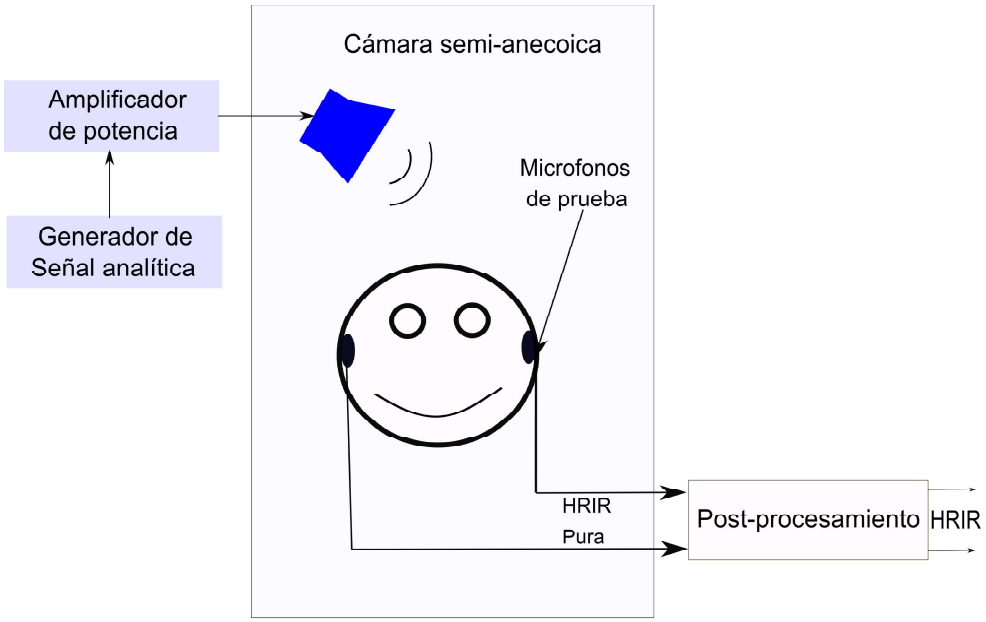

Figura 1.6: Diagrama de bloques para la medición de las HRIR. Adaptado de [1].

procesar utilizando micrófonos ubicados cerca del canal auditivo [1], como muestra la Figura 1.6.

Este procedimiento se repite para cada dirección deseada. Una alternativa a la descripción dada anteriormente es que los altavoces estén en una posición fija y la persona gire para registrar todas las posiciones.

A continuación se detallarán ciertas consideraciones a tener en cuenta el momento de realizar las mediciones de las HRTFs.

Señal analítica de entrada: Se han empleado varios tipos de señales para la obtención de HRIRs, como: señales determinísticas (e.g. impulso [16] , barridos sinusoidales [17]), señales aleatorias (e.g. ruido blanco, ruido rosa) y señales de ruido pseudoaleatorias. Al momento de realizar mediciones de esta naturaleza, un problema a tenerse en cuenta siempre, es la presencia de ruido en los sistemas de medición, por lo tanto, es necesario que las características de la señal de entrada ayuden a mejorar la relación señal a ruido (SNR). Si bien una primera alternativa consistiría en aumentar el nivel de potencia de la señal de entrada, un aumento excesivo de la potencia podría causar distorsión en los sistemas electroacústicos como altavoces y amplificadores. Por lo tanto, para aumentar el nivel de potencia sin violar la linealidad de los sistemas electroacústicos, una señal de entrada ideal debería tener un bajo factor de cresta (relación de valores pico a valor efectivo) [18].

Posición de la sonda de micrófono: La Figura 1.7 muestra la ubicación del micrófono, el

cual se coloca comúnmente en un punto cercano a la entrada del canal auditivo bloqueado. Esta posición del micrófono introducida por [2] ha sido ampliamente utilizada por su comodidad y seguridad. Es posibles usar otras posiciones para la ubicación de los micrófonos, por ejemplo, cerca de la entrada del canal auditivo sin bloquearlos o dentro del canal auditivo cerca del tímpano [1].

Figura 1.7: Posición de sonda de micrófono [1].

Individuos: Los HRIRs (y por lo tanto las HRTFs) son funciones que dependen de las características anatómicas de cada individuo, esto hace que sea difícil tener un conjunto de HRIRs generales. Por esta razón, las HRIRs se miden para varios sujetos. Dado que es difícil mantener a las personas quietas durante largos períodos de tiempo, ya que tienden a hacer pequeños movimientos de la cabeza y el cuerpo, las cabezas artificiales y los torsos como el KEMAR(ver Figura 1.8) (*Knowles Electronics Manikin For Acoustics Research* [19]) se utilizan comúnmente. KEMAR se diseñó basándose en las dimensiones promedio de las características anatómicas de acuerdo con los requisitos de ANSI S3.36 / ASA58- 1985 y IEC 60959: 1990 [20].

Cámara Semi-anecoica: Es común que las mediciones de HRIR se realicen en una cámara anecoica para eliminar posibles reflejos del medio ambiente. Sin embargo, debido a las dificultades técnicas y los altos costos de una cámara anecoica, las mediciones HRIR también se han realizado en cámaras semi-anecoicas [22]. En cámaras semi-anecoicas, el tiempo de llegada de las ondas reflectadas se puede controlar para que lleguen después de la duración de las HRIRs (normalmente unos pocos milisegundos). Esto se puede lograr colocando material absorbente acústico en la habitación [1]. En estas condiciones, se aplica una

Figura 1.8: Maniquí KEMAR. Fuente [21]

ventana de tiempo en las HRIRs sin procesar para eliminar las reflexiones no deseadas [22].

1.3.2.3. Principales características de las HRTFs

Características en el dominio del tiempo de las HRIR

Dentro del dominio del tiempo la pista más relevante a considerar es la ITD. Aunque hay varios métodos para determinar la ITD, existe un enfoque directo para su estimación, que es calcular la diferencia entre los retrasos de la HRIR del el oído izquierdo y la HRIR del oído derecho [1]. Como muestra la Figure 1.9. Por lo tanto, la ITD se define como:

$$ITD(\theta, \phi) = t_L - t_R \tag{1.4}$$

donde, t_L y t_R son los retrasos de inicio para el oído izquierdo y derecho respectivamente.

Características en el dominio de la frecuencia de las HRTFs

Podemos analizar las características de frecuencia de las HRTFs fijándonos en la Figura 1.10 que muestra las magnitudes de las HRTFs para cuatro ángulos de azimut en el plano horizontal, donde podemos observar que, en la Figura 1.10a ($\theta = 0°$) las HRTFs izquierda y derecha son similares, las pequeñas diferencias entre ellas se deben al hecho de que la cabeza no es perfectamente simétrica. En frecuencias bajas (por debajo de la banda de 400 -500 Hz), el efecto de atenuación de la cabeza es despreciable (esto debido a la longitud de onda que presentan las frecuencias en esta banda). A medida que la frecuencia aumenta por encima de 1,5 KHz, las diferencias de nivel interaural (ILD) se hacen más evidentes. Por lo tanto, la diferencia entre las magnitudes del oído izquierdo con el oído derecho, va en aumento a partir de esta frecuencia. Vale la pena señalar que esta diferencia aumenta con el ángulo de azimut. Esta diferencia es la ILD, y está definida como:

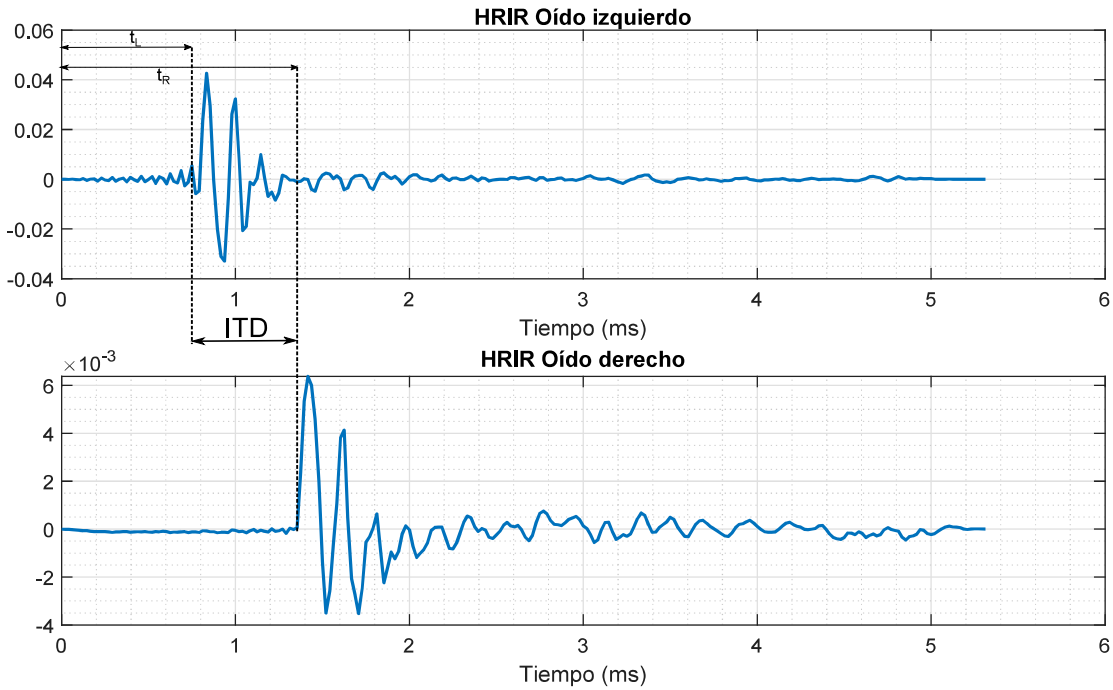

Figura 1.9: Representación de la ITD de sujeto NH2 de la base de datos ARI para $\theta = 80°$ y $\phi = 0°$.

$$ILD(\theta, \phi) = 20log_{10}\frac{|H_R(\theta, \phi, f)|}{|H_L(\theta, \phi, f)|} \tag{1.5}$$

Finalmente, dos características importantes a destacar son que, en altas frecuencias, desde la banda de 5 a 6 kHz, las magnitudes de las HRTF varían de manera compleja y presentan varios picos y muescas. Así mismo, se observa que existe un pico cercano a 4 kHz debido a la resonancia del canal auditivo [1].

Ahora bien, si se busca implementar un sistema de sonido 3D un problema importante a considerar, es el hecho de que las características espectrales de las HRTFs difieren ampliamente de un individuo a otro.

La Figura 1.11 muestra las magnitudes de las HRTFs en decibelios del oído izquierdo para cuatro sujetos diferentes, en una ubicación específica. La variabilidad entre los sujetos es especialmente notable para las frecuencias altas (es decir, frecuencias mayores a 4 KHz) donde las pistas monoaurales introducidas por el pabellón auditivo son más prominentes [23]. Varios estudios muestran que existe disminución en la precisión con la cual se localiza una fuente se sonido debido a las HRTFs no individualizadas [24], [25]. Esto podría provocar en el oyente: confusiones frente-atrás y arriba-abajo, así como también, una exteriorización de sonido deficiente (es decir, el sujeto percibe el sonido dentro de su cabeza) y percepción incorrecta de elevación.

Por lo tanto, resulta necesario personalizar las HRTFs para garantizar una percepción del

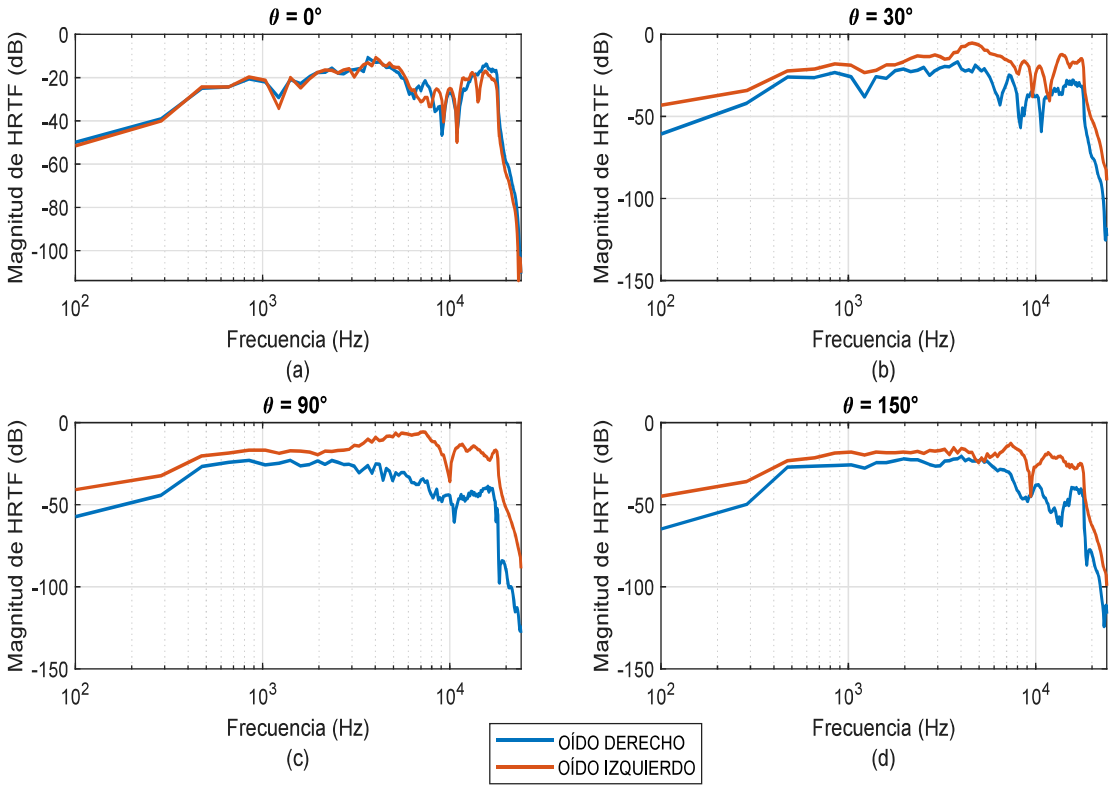

Figura 1.10: Magnitudes de HRTFs de la base de datos ARI para varios azimuts en el plano horizontal.

sonido 3D de alta calidad. El enfoque más preciso para personalizar las HRTFs es a través de mediciones directas.

1.3.2.4. Bases de datos de las HRTFs

Varios equipos de investigación han logrado construir bases de datos de HRTFs disponibles públicamente. Con el fin de facilitar el intercambio de datos entre investigadores y usuarios, la Sociedad de Ingenieros de Audio (*Audio Engineering Society*, AES) ha estandarizado el Formato Orientado Espacialmente para la Acústica (*Spatially Oriented Format for Acoustics*, SOFA).

SOFA es un formato de archivo para almacenar e intercambiar datos acústicos, como las HRTFs. Actualmente, las bases de datos HRTF más populares están disponibles públicamente para descarga en formato SOFA.

Por ejemplo, la base de datos ARI con cerca de 100 sujetos estudiados, cuenta con 1550 posiciones para cada oyente, incluido el espacio azimutal completo de $0°$ a $360°$ y de elevación de $-30°$ a $+80°$, con resolución de 2,5° para el azimut y $5°$ para la elevación [26]. Otra base de datos popular es CIPIC [22] que cuenta con mediciones para 43 sujetos, en la cual, a diferencia de ARI se realizaron 1250 mediciones en coordenadas interaurales para 25 ángulos de azimut distribuidos no uniformemente con una resolución máxima de $5°$ para ubicaciones cercanas al plano medio y 50 mediciones de elevación con resolución 5.625°

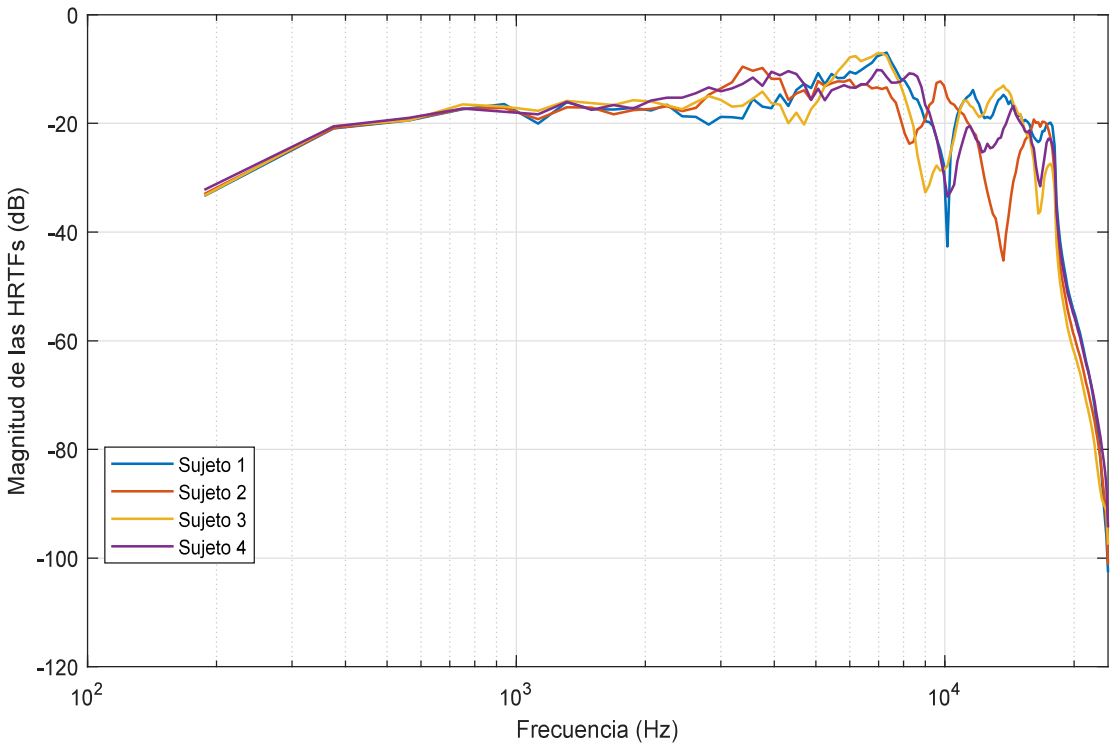

Figura 1.11: HRTFs para el oído izquierdo en una ubicación específica para cuatro sujetos diferentes.

en el rango de $-45°$ a 230.625°.

Formato SOFA

SOFA apunta a representar las HRTFs de una manera general, medidas con una conjunto de micrófonos que representan al oyente y excitada por una matriz de altavoces que representan a las fuentes de sonido [4].

En el formato SOFA los datos y los metadatos se describen como arreglos cuyas dimensiones vienen definidas en el formato: $[A_1 A_2 ... A_i]$ (ver Tabla 1.2), donde A_i representa la longitud de la i-ésima dimensión del arreglo.

Las especificaciones de SOFA buscan simplificar el intercambio de información acústica a través de interfaces de programación para Matlab, Octave y C ++. SOFA además proporciona estándares para una descripción coherente de las configuraciones de medición para futuras bases de datos de HRTFs [4].

Dentro del archivo SOFA se tiene varios parámetros, sin embargo, el campo Data es el más importante en el archivo, ya que este almacena los atributos: IR, Delay, $\mathrm{SamplingRate}$ y $\mathrm{SamplingRate_Units}$. Otro parámetro a tener presente es el campo $\mathrm{SourcePosition}$, que es una matriz que contiene las posiciones de las fuentes de sonido en coordenadas esféricas. El campo Data, representa una descripción numérica de los sistemas acústicos (e.g el sistema auditivo humano) y consiste en un arreglo multidimensional de tamaño arbitrario. Los datos

Tabla 1.1: Dimensiones definidas en SOFA [4].

Dimensión	Valor	Descripción
I	1	Define un valor escalar.
M	ilimitado	Número de mediciones. Debe ser un entero mayor que cero.
R	ilimitado	Número de receptores. Debe ser un entero mayor que cero.
E	ilimitado	Número de emisores. Debe ser un entero mayor que cero
N	ilimitado	El número de muestras de datos que describen una medición. Debe ser un número entero mayor que cero.
C	3	Dimensión de las coordenadas. Siempre es un valor igual a 3.

Tabla 1.2: Atributos del campo Data considerados en el formato SOFA [4].

Nombre	Valor por defecto	Dimensiones	Tipo	Descripción
Data.IR	0	[M R N]	double	Respuesta impulsiva a lo largo del eje del tiempo.
Data.Delay	0	[I R]	double	Tiempo de llegada en muestras para cada receptor (*Time of Arrival*, TOA).
Data.SamplingRate	48000	[I], [M]	double	Frecuencia de muestreo.
Data.SamplingRate_Units	hertz	N/A	string	Unidades de muestreo de la frecuencia de muestreo.

almacenados en este formato tienen la flexibilidad de estar en el dominio que mejor se adapta al sistema de medición [4]. Los atributos del campo Data son descritos en la Tabla 1.2. A continuación se presenta un ejemplo del uso de dimensiones en el formato SOFA.

Supongamos una base de datos que consta de mil mediciones, es decir, M = 1000, obtenidas para 1000 posiciones diferentes. Si usamos dos micrófonos, es decir, dos IR por medición (i.e. R = 2), y una tasa de muestreo de 48 KHz, tendremos que el campo $\mathrm{SourcePosition}$ es una matriz de dimensiones [M C]. Además, si se asume una sola posición de escucha, es decir que el oyente no cambia de posición durante la medición de las HRTFs, tendremos que $\mathrm{Data.IR}$, $\mathrm{SourcePosition}$ y $\mathrm{ListenerPosition}$ (i.e. posición del oyente durante las mediciones) serán de dimensión [1000 2 N], [1000 3] y [1 3], respectivamente.

1.3.2.5. Aproximación de fase mínima de las HRTFs

Las HRTF son funciones de valores complejos, por lo tanto, incluyen componentes de magnitud y fase. De forma general, cualquier función de transferencia y, por lo tanto, cualquier HRTF se puede descomponer [27] como el producto de su función de fase mínima $H_{min}(\theta, \phi, f)$, una función pasa-todo (i.e. *all-pass function*) $exp[j\Psi_{all}(\theta, \phi, f)]$ de ganancia unitaria y una función de fase lineal $exp[-2j\pi fT(\theta, \phi)]$, teniendo asi que :

$$H(\theta, \phi, f) = H_{min}(\theta, \phi, f) \cdot exp[j\Psi_{all}(\theta, \phi, f)] \cdot exp[-2j\pi fT(\theta, \phi)] \tag{1.6}$$

donde, $T(\theta, \phi)$, es el retardo de tiempo sin procesar causado por la propagación de la fuente de sonido al oído y Ψ representa la fase de la señal. Por otro lado, sabemos que la magnitud de una función de transferencia y su magnitud de fase mínima correspondiente son iguales [27], es decir: $|H| = |H_{min}|$, por lo tanto tenemos que:

$$H_{min}(\theta, \phi, f) = |H(\theta, \phi, f)| \cdot exp[j\Psi_{min}(\theta, \phi, f)] \tag{1.7}$$

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

$$\Psi_{min}(\theta, \phi, f) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\ln |H(\theta, \phi, x)|}{f - x} \cdot dx \tag{1.8}$$

Ahora, la Ecuación 1.6 muestra que de forma general, la fase de las HRTFs comprende una fase mínima, una fase de la componente pasa-todo y una fase lineal. Entonces podemos expresar la fase de las HRTFs como:

$$\Psi_{min}(\theta, \phi, f) = \Psi_{min}(\theta, \phi, f) + \Psi_{all}(\theta, \phi, f) - 2\pi f T(\theta, \phi) \tag{1.9}$$

Entonces si consideramos que la componente de fase de la función pasa-todo (all-pass) es despreciable, una HRTF puede aproximarse como:

$$H(\theta, \phi, f) \approx |H(\theta, \phi, f)| \cdot exp[-j2\pi f T(\theta, \phi, f)] \tag{1.10}$$

La Ecuación 1.10 es conocida como la aproximación de fase mínima de las HRTF, en la que una HRTF se aproxima por su función de fase mínima en cascada con una fase lineal $T(\theta, \phi, f)$ (i.e. un retardo puro que corresponde a la ITD). En este punto cabe mencionar que, el retraso se colocaría solo en la HRTF (izquierda o derecha) que garantiza que el tiempo de inicio sea siempre positivo.
La valides del modelo de fase mínima de las HRTFs ha sido demostrada por los estudios realizados por Kistler,D. [28] y Kulkarni,S. [29]. Esta aproximación presenta una gran ventaja ya que permite procesar la magnitud y fase de las HRTF de manera independiente. Luego, la HRTF compleja se puede reconstruir utilizando la Ecuación 1.10. En este proyecto utilizamos esta aproximación, por lo tanto, de aquí en adelante nos enfocaremos en la magnitud únicamente.

1.3.3. ANÁLISIS DE COMPONENTES PRINCIPALES

El análisis de componentes principales o PCA por sus siglas en ingles (*Principal Component Analysis*) es una alternativa para la resolución de un problema de aprendizaje no supervisado conocido como reducción de dimensionalidad [3]. Esta técnica suele ser utiliza para revelar relaciones entre las variables en un conjunto de datos y de esta manera reducir un

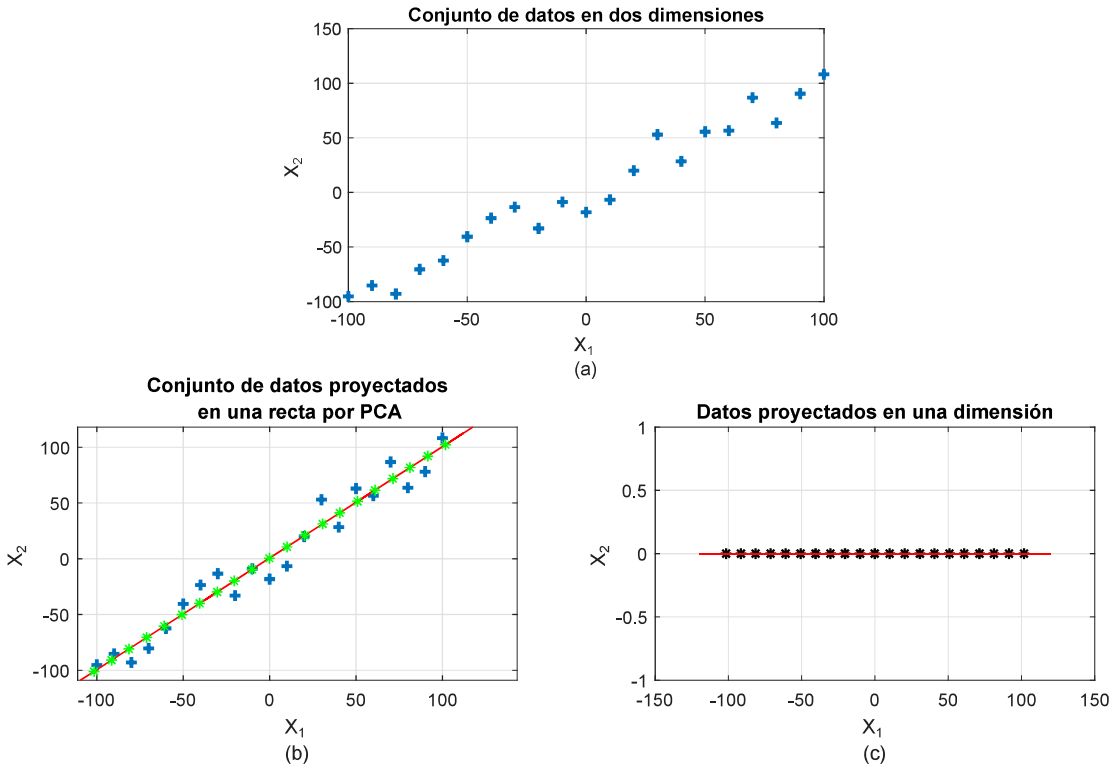

Figura 1.12: Ejemplo de aplicación de PCA en datos en $\mathbb{R}^2$.

conjunto de datos de alta dimensión en un espacio más significativo de baja dimensión [30]. Existen varias razones por las cuales uno podría querer hacer una reducción de dimensionalidad. Sin embargo, estas se pueden resumir en dos grandes aplicaciones. Una de ellas es la compresión de datos, ya que la compresión no sólo nos permite reducir la cantidad de datos y por lo tanto utilizar menos memoria de la computadora o espacio en el disco, también nos permitirá acelerar algoritmos de aprendizaje. Por otro parte, PCA al poder reducir la dimensionalidad también es aplicado para revelar características de los datos que no pueden ser apreciadas en datos de alta dimensión. En este caso PCA es considerada una técnica de visualización de datos [31].

1.3.3.1. Formulación del problema de PCA

Suponga un conjunto de datos en $\mathbb{R}^2$ como el que se muestra en la Figura 1.12.a. Nuestro objetivo será reducir la dimensión de los datos, de dos dimensiones a una sola dimensión. En otras palabras, buscamos una linea sobre la cual proyectar los datos que están en dos dimensiones. Esto con el fin de representar nuestros datos en una única variable, y así reducir el espacio de almacenamiento requerido.

Ahora bien, la línea roja mostrada en la Figura 1.12.b podría ser una buena opción sobre la cual proyectar nuestros datos, dado que, si observamos los datos originales (i.e. puntos azules), y su correspondiente proyección sobre la línea roja (i.e. punto verdes), encontramos que la distancia entre cada punto y su versión proyectada es muy pequeña.

El ejemplo anterior corresponde a una reducción de dimensionalidad, en el cual se ha tomado un conjunto de datos en $\mathbb{R}^2$ y se lo ha reducido a un conjunto en una dimensión, como muestra la Figura 1.12.c.

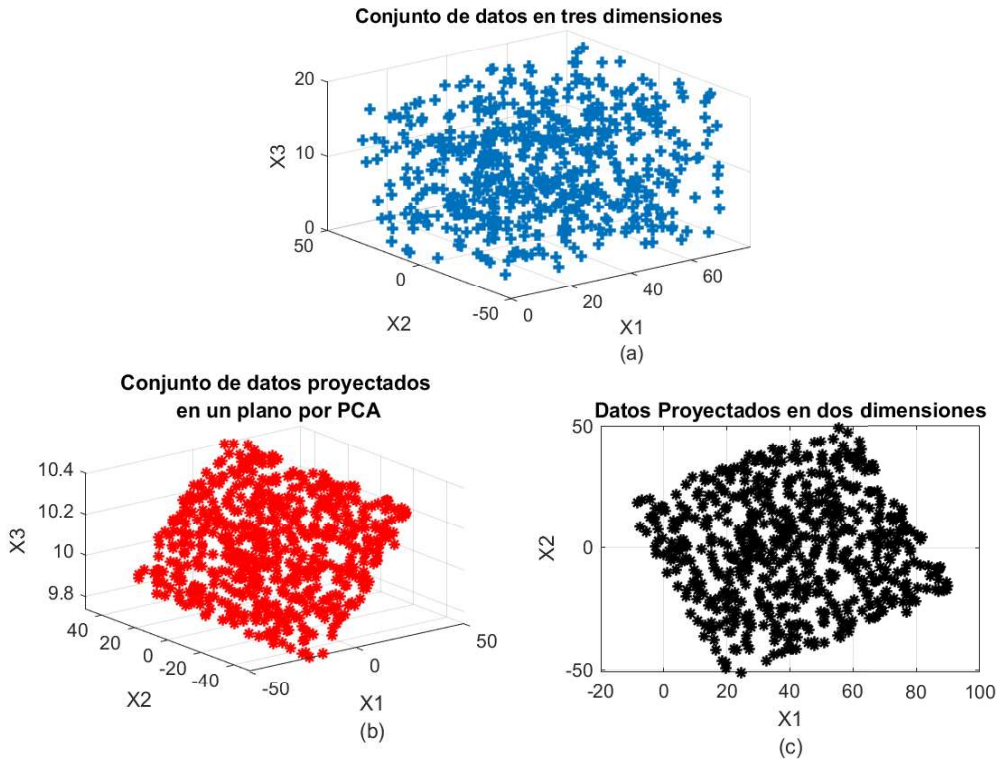

Figura 1.13: Ejemplo de datos en $\mathbb{R}^3$ proyectados en $\mathbb{R}^3$ usando PCA.

Este ejemplo puede ser llevado a conjuntos de datos de dimensión arbitraria. Por ejemplo, la Figura 1.13, muestra la reducción de dimensionaldad de un conjunto de datos en $\mathbb{R}^3$, a un conjunto de datos en $\mathbb{R}^2$. En este punto es importante destacar que, el caso más típico de reducción de dimensionalidad, sería tener datos de dimensión muy alta (e.g. 1000 dimensiones) que quisiéramos reducir a, por ejemplo, un centenar de dimensiones. Debido a las limitaciones visuales para representar datos en alta dimensión se ha optado por utilizar ejemplos intuitivos(e.g datos en $\mathbb{R}^2$ y $\mathbb{R}^3$) para ilustrar la idea de reducción de dimensionalidad.

Tabla 1.3: Conjunto de datos originales o de alta dimensión. Adaptado de [30]

País	$\mathbf{x}_1$: PIB (Trillones de Dolares americanos)	$\mathbf{x}_2$: PIB per Capital	$\mathbf{x}_3$: Índice de desarrollo humano	$\mathbf{x}_4$: Esperanza de vida	$\mathbf{x}_5$: Índice de pobreza	...	$\mathbf{x}_{50}$
Canada	1.577	39.71	0.908	80.7	32.6	...	...
China	5.878	7.54	0.687	73	49.6	...	...
India	1632	3.41	0.547	64.7	36.8	...	...
Rusia	1.48	19.84	0.755	65.5	39.9	...	...
Singapur	0.223	56.69	0.866	80	42.5	...	..
USA	14.527	46.86	0.91	78.3	40.8	...	...
...	...	...	...	...	...	...	...

Ahora bien, PCA también puede ser usada como técnica de visualización para revelar características relevantes de los datos que no se aprecian en alta dimensionalidad. Por ejemplo, supongamos que se han recopilado un gran conjunto de datos de varias fuentes estadísticas sobre diferentes países alrededor del mundo. Así, tenemos que el primer atributo, $\mathbf{x}_1$ es el PIB del país, o el Producto Interno Bruto, $\mathbf{x}_2$ es el PIB per cápita, lo que significa el PIB por Persona, $\mathbf{x}_3$ es el Índice de Desarrollo Humano, $\mathbf{x}_4$ la Esperanza de vida, $\mathbf{x}_5$ Índice de pobreza, y así sucesivamente [30].

Sin embargo, este conjunto de datos puede crecer aun más (para el caso de este ejemplo un total de 50 dimensiones $\mathbf{x}_{50}$, como muestra la Tabla 1.3), haciendo que la visualización de los datos sea cada vez mas compleja. Por lo tanto, nuestro objetivo será representar a cada país ya no como un vector de 50 dimensiones, en el cual cada dimensión representa un atributo o variable, sino comprimir la información de estos 50 atributos a un número reducido de ellos.

Si ese es el caso, en la Tabla 1.4 se observa como un sólo par de variables, $\mathbf{z}_1$ y $\mathbf{z}_2$, resumen la información de los atributos presentes en la Tabla 1.3, con lo que se ha podido reducir los datos de 50 dimensiones a 2 dimensiones. Ahora bien, si graficamos los datos de la Tabla 1.4 en la Figura 1.14, cada país está representado por una coordenada $\mathbf{z}^{(i)} \in \mathbb{R}^{(2)}$. A continuación intentaremos dar una interpretación a las nuevas variables obtenidas.

El eje horizontal $\mathbf{z}_1$ corresponde aproximadamente al tamaño total del país, o la actividad económica general de un país (i.e. PIB), mientras que el eje vertical $\mathbf{z}_2$ en nuestros datos corresponde al PIB por persona.

Entonces, tomando en cuantas estas consideraciones, observamos en la Figura 1.14 que Estadios Unidos es un país con una gran actividad económica y presenta además un PIB por persona alto. Por otra parte, Singapur presenta un PIB por persona similar la de Estado Unidos, sin embargo, su actividad económica general es muy pequeña. Así hemos podido representar la información de la Tabla 1.3, con únicamente las variables $\mathbf{z}_1$ y $\mathbf{z}_2$.

Tabla 1.4: Conjunto de datos en baja dimensión. Adaptado de [30]

Pais	$\mathbf{z}_1$	$\mathbf{z}_2$
Canada	1.6	1.2
China	1.7	0.3
India	1.6	0.2
Rusia	1.4	0.5
Singapur	0.5	1.7
USA	2	1.5
...	...	...

En conclusión, hasta el momento hemos analizado a través de ejemplos intuitivos las dos grandes aplicaciones de PCA. Por una parte, la reducción de dimensionalidad es una herramienta útil al momento de comprimir datos y, por otro lado, la visualización de datos nos permite encontrar las características más relevantes en un conjunto de datos con múltiples variables.

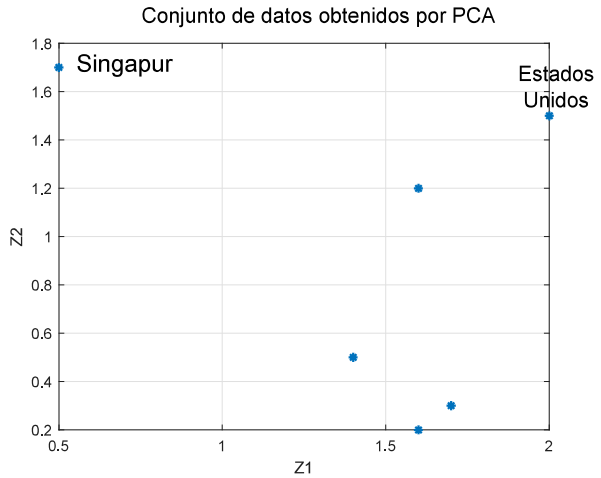

Figura 1.14: Gráfica de la tabla 1.4.

1.3.3.2. Algoritmo de PCA

Previa a la aplicación de PCA, se debe realizar una etapa de pre-procesamiento para un conjunto de datos de entrenamiento representado por la matriz **X** de tamaño $m \times n$, donde cada columna corresponderá a una variable, y cada fila a una muestra u observación, es decir, tendremos un total de m observaciones y n de variables.

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & X_{1,2} & X_{1,3} & ... & X_{1,j} & ... & X_{1,n} \\ X_{2,1} & X_{2,2} & X_{2,3} & ... & X_{2,j} & ... & X_{1,n} \\ X_{3,1} & X_{3,2} & X_{3,3} & ... & X_{3,j} & ... & X_{1,n} \\ ... & ... & ... & ... & ... & ... & .... \\ X_{m,1} & X_{m,2} & X_{m,3} & ... & X_{m,j} & ... & X_{m,n} \end{pmatrix}$$

Se tendrá entonces que $\mathbf{x}_j$ es un vector columna, que representa a una variable j-ésima de los datos de entrenamiento, a la cual se la normalizará para que la media sea cero y la escala sea similar a través de un escalamiento de variable. Además $x^{(i)}$ es un vector fila que representa la i-ésima observación de los datos de entrenamiento.

Entonces, para la normalización de la media, primero obtenemos la media de cada variable (ver Ecuación 1.11) y a continuación, reemplazamos cada variable $\mathbf{x}_j$ con $\mathbf{x}_j$ menos su media, (i.e. $\mathbf{x}_j = \mathbf{x}_j - \mu_j$). Esto hace que cada variable ahora tenga una media de cero.

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} X_{i,j} \tag{1.11}$$

El escalamiento es necesario ya que las diferentes variables pueden tener escalas muy distintas. Por ejemplo, si $\mathbf{x}_1$ representa el tamaño de una casa, y $\mathbf{x}_2$ es el número de habitaciones, es necesario escalar cada variable para tener una gama de valores comparables. Así, tendríamos que

$$\mathbf{x}_j = \frac{\mathbf{x}_j - u_j}{s_j}, \tag{1.12}$$

donde, s_j es cierta medida de la gama de valores de la variable j (i.e. la variabilidad de $\mathbf{x}_j$. Por ejemplo, s_j podría ser el valor máximo menos el mínimo, o la desviación estándar de la variable j [30].

Una vez realizado el pre-procesamiento, se procederá a reducir los datos, de n dimensiones a k dimensiones. Con este objetivo, lo primero a hacer en el algoritmo de PCA es calcular la matriz de covarianza denotada comúnmente por medio de la letra griega □ y definida como

$$\square = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}^{(i)})(\mathbf{x}^{(i)})^T = \frac{1}{m}\mathbf{X}^T\mathbf{X}, \tag{1.13}$$

donde, $\mathbf{x}^{(i)}$ corresponde a la i-ésima observación de la matriz **X** (i.e. la i-ésima fila). Con respecto a la matriz de covarianza, es importante mencionar que:

- La matriz de covarianza será una matriz cuadrada de tamaño $n \times n$
- Los valores de la diagonal de la matriz □ (i.e. $\Sigma_{j,j}$) representan la varianza de j-ésima variable.
- Los valores fuera de la diagonal representan los valores de la covarianza entre las variables i y j.

A continuación, procedemos a realizar la Descomposición en Valores Singulares (*Singular Value Decomposition*, SVD) de la matriz de covarianza, de la cual se obtiene una matriz **V** cuyas columnas contienen los vectores propios de la matriz □. Cada vector propio se corresponderá con un valor propio, cuya magnitud indica cuánta variabilidad de los datos se explica por el mismo.

Ahora bien, dado que **V** es una matriz cuadrada de tamaño $n \times n$ cuyos vectores columna $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ representan los vectores propios de la matriz □, si queremos reducir los datos de n dimensiones hasta k dimensiones, basta con tomar las primeras k primeras columnas de a matriz **V** [30]. Se escogen las primeras k columnas ya que el método SVD retorna los vectores propios en orden descendente con base en la variabilidad de cada uno. Y puesto que la variabilidad es un indicador de cuánta información retiene cada vector propio, al tomar las primeras k columnas de la matriz **V**, garantizamos que, a pesar de la reducción de dimensionalidad, se conserve una gran cantidad de la información original [32].

Las primeras k columnas de **V** pueden ser agrupadas como una nueva matriz $\mathbf{V}_{\textbf{reducida}}$ de tamaño $n \times k$, que es conocida como matriz de rotación, la misma que define un nuevo espacio vectorial, ya que está compuesta íntegramente por vectores singulares [32].

$$\mathbf{V}_{reducida} = \begin{pmatrix} V_{1,1} & V_{1,2} & V_{1,3} & V_{1,k} \\ V_{2,1} & V_{2,2} & V_{2,3} & V_{1,k} \\ V_{3,1} & V_{3,2} & V_{3,3} & V_{1,k} \\ ... & ... & ... & ... \\ V_{n,1} & V_{n,2} & V_{n,3} & V_{n,k} \end{pmatrix}$$

Finalmente, procedemos a proyectar los datos originales sobre el nuevo espacio vectorial de baja dimensión a través de la multiplicación vectorial

$$\mathbf{z}^{(i)} = \mathbf{x}^{(i)}\mathbf{V}_{reducida}, \tag{1.14}$$

donde el vector fila $\mathbf{z}^{(i)}$ corresponde a la proyección en el nuevo espacio vectorial reducido de la i-ésima observación de la matriz (X). Los vectores $\mathbf{z}^{(i)}$ pueden agruparse dentro de la matriz **Z** que tendrá dimensiones $m \times k$ (i.e. m observaciones de k dimensiones).
Esta información reorientada corresponde a las nuevas coordenadas de los datos originales proyectados en baja dimensión, y se la conoce como *score* o peso.

1.3.3.3. Reconstrucción a partir de los datos de baja dimensionalidad

Hasta este punto hemos definido a PCA como un algoritmo de compresión, en el que por ejemplo, podríamos tener datos de 1000 dimensiones y comprimirlos a un vector de variables en 100 dimensiones. Por lo tanto, es posible pensar que debe haber una manera en la cual a partir de la representación comprimida de datos, se pueda volver a una aproximación de los datos en la dimensión original (i.e. alta dimensionalidad). A este proceso se lo conoce como reconstrucción.

Considere el siguiente ejemplo. Para un conjunto de datos $\mathbf{X} \in \mathbb{R}^2$, si tomamos dos observaciones en 2D $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ y los proyectamos en una recta, tendremos que los pares ordenados anteriormente mencionados pueden ser representados por, $\mathbf{z}^{(1)}$ y $\mathbf{z}^{(2)}$ respectivamente, que son vectores 1D [30]. La Figura 1.15.a muestra los observaciones originales (i.e. puntos verdes), mientras la Figura 1.15.b muestra la proyección de los datos en una dimensión (i,e puntos negros) usando PCA. Así hemos reducido la dimensionalidad de los datos originales de dos dimensiones a una dimensión.

Ahora que hemos aplicado PCA, nuestro objetivo será llevar a cabo la reconstrucción aproximada del conjunto de datos **X**, para lo cual, partiremos del vector **z**, que contiene los valores correspondientes a las proyecciones de **X** sobre una recta (i.e. $\mathbf{z}^{(i)}$). Podemos llevar a cabo esta reconstrucción del conjunto de datos **X** a través de

$$\mathbf{x}^{(i)}_{aprox} = \mathbf{z}^{(i)}[\mathbf{V}_{reducido}]^T + \mu, \tag{1.15}$$

donde, μ es un vector de tamaño $1 \times n$ que contiene la media de cada variable (i.e. de cada columna) de la matriz **X**. Esta media se calcula como parte el proceso de PCA al normalizar los datos de entrenamiento.

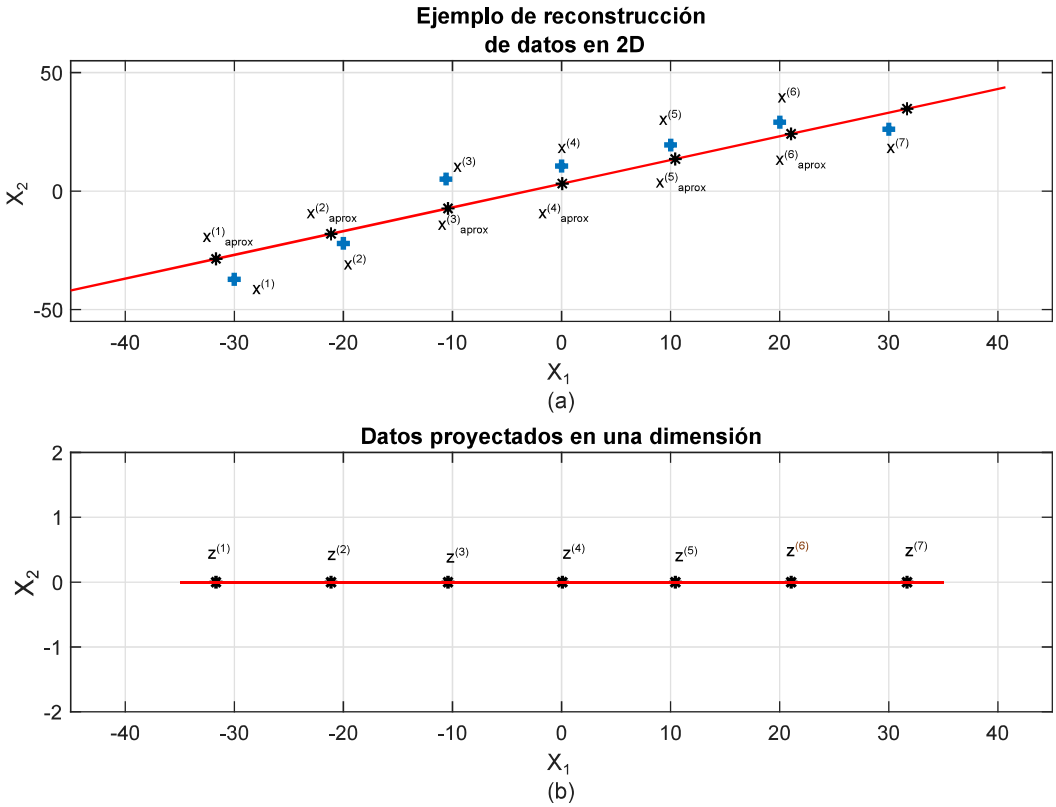

Figura 1.15: Ejemplo de reconstrucción.

Lo que se obtiene de la Ecuación 1.15 es una aproximación de los valores originales de **X**, que de forma gráfica se los observa en la Figura 1.15.a, como los puntos azules que se encuentran sobre recta roja [33]. En otras palabras, el resultado de la Ecuación 1.15 para las muestras $\mathbf{z}^{(1)}$ y $\mathbf{z}^{(2)}$, serán respectivamente $\mathbf{x}^{(1)}_{aprox}$ y $\mathbf{x}^{(2)}_{aprox}$. Si bien, se puede observar que existe un cierto grado de error entre los datos originales y sus aproximaciones, este error puede ser modificado con base en el número de componentes principales que se seleccionen.

1.3.3.4. Selección del número de componentes principales

Al presentar el algoritmo de PCA tomamos un conjunto de datos de entrenamiento de n dimensiones y lo reducimos a cierta representación de k dimensiones. Este número k es el parámetro mas importante del algoritmo de PCA, ya que este representa el número de componentes principales o número de componentes retenidas.

Para seleccionar el número de componentes principales, se deben considerar indicadores como, el error cuadrático medio (*Mean Squared Error*, MSE), definido como

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m} ||\mathbf{x}^{(i)} - \mathbf{x}^{(i)}_{aprox}||^2 \tag{1.16}$$

para el cual nuestro objetivo será que su valor sea lo más pequeño posible. En otras palabras, este valor estima la diferencia media entre los datos originales $\mathbf{x}^{(i)}$ y su versión reconstruida, $\mathbf{x}^{(i)}_{aprox}$.

Por otra parte, podemos también estimar la variación media de las observaciones usando la Ecuación 1.17 [30]

$$\frac{1}{m}\sum_{i=1}^{m}||\mathbf{x}^{(i)}||^2, \tag{1.17}$$

Ahora bien, con base en los indicadores mencionados anteriormente, se procede a escogerá el valor de k. Una regla general bastante común, es elegir valores que hagan que la relación entre la Ecuación 1.16 y la Ecuación 1.17 sea menor que 0.01. Otros valores comunes suelen ser 0.05 y 0.1 [1].

En otras palabras, el objetivo principal al momento de escoger el número de componentes principales es que el MSE, es decir, la distancia promedio entre **X** y sus proyecciones, dividida entre la variación media de las observaciones, sea lo menor posible, i.e.,

$$\frac{\frac{1}{m}\sum_{i=1}^{m}||\mathbf{x}^{(i)}-\mathbf{x}^{(i)}_{aprox}||^2}{\frac{1}{m}\sum_{i=1}^{m}||\mathbf{x}^{(i)}||^2} \tag{1.18}$$

Así, por ejemplo, si el valor de la Ecuación 1.18 es 0.01 para un cierto k, esto representará que con k componentes seleccionados se ha perdido el 1 % de información, o dicho de otra forma, se tiene el 99 % de varianza retenida con k componentes principales.

A modo de resumen podemos decir que, para seleccionar el número de componentes principales, se debe aplicar el algoritmo de PCA a un conjunto de datos **X** y seleccionar por ejemplo $k = 3$ componentes. A continuación se debe calcular $\mathbf{X}_{apro}$ usando la Ecuación 1.15, para en base a esta aplicar la Ecuación 1.18 y así determinar si cumple con el porcentaje de varianza retenida deseada. Caso contrario se debe repetir el procedimiento para un número diferente de k.

Este técnica puede ser bastante incomoda ya que requiere ejecutar el algoritmo de PCA varias veces. Afortunadamente, este procedimiento se simplifica cuando se usa SVD como método para obtener los vectores y valores singulares de la matriz de covarianza □. El método SVD nos devuelve de hecho una matriz diagonal que denominaremos **S** la cual es la matriz de covarianza de los datos proyectados **Z**. La diagonal de **S** (i.e. $S_{1,1}$, $S_{2,2}$, $S_{3,3}$, ... $S_{n,n}$) son los únicos valores diferentes de 0 en la matriz **S** y representan en orden decreciente los valores de varianza para cada componente principal respectivamente [32]. Entonces, el cálculo de k se simplifica tal que, se determinará el número de componentes principales con la siguiente expresión [30]

$$\frac{\sum_{i=1}^{k} S_{i,i}}{\sum_{i=1}^{n} S_{i,i}}, \tag{1.19}$$

en la cual probaremos variando el valor de k. Si el valor arrojado por la Ecuación 1.19 con

por ejemplo, $k = 10$ es mayor o igual a 0.95, esto asegura que se a retenido el 95 % de la varianza, es decir, se a conservado el 95 % de la información original. La Ecuación 1.19 es conocida como la varianza retenida acumulada.

Este procedimiento resulta ser mucho más eficiente y permite seleccionar el valor de k sin necesidad de ejecutar el algoritmo de PCA desde cero una y otra vez.

1.3.3.5. Análisis de componentes principales para HRTFs

Como ya hemos visto hasta ahora el análisis de componentes principales es una poderosa técnica estadística que se utiliza para reducir la dimensionalidad en un conjunto de datos correlacionados y simplificarlo mientras conserva la información de los datos originales. Esto nos permite presentar de forma compacta las funciones HRTFs, manteniendo la información más relevante de las mismas. Para el caso de las HRTFs las variables descritas en la Sección 1.3.3.2, corresponden a las componentes de frecuencia de cada HRTF, mientras que las observaciones corresponden a las diferentes ubicaciones de la fuente de sonido. Generalmente PCA se aplica antes de procesos de personalizaron o interpolación [34]. Es por esto que, PCA se convierte en una herramienta importante al momento de analizar las características en frecuencias de las las funciones HRTFs [1].

La descomposición de las funciones HRTFs usando PCA se muestra a continuación como

$$\mathbf{H}(\theta, \phi, f, s) - H_{av}(f) = \sum_{q} \mathbf{w}_q(\theta, \phi, s) \cdot \mathbf{d}_q(f) \tag{1.20}$$

donde, los parámetros, θ y ϕ denotan la ubicación de la fuente de sonido, f indica una frecuencia discreta y s identifica al sujeto. A demás, $\mathbf{d}_q(f)$ es una serie de funciones base (i.e. los vectores base de la matriz **V**) que dependen de la frecuencia, $\mathbf{w}_q(\theta, \phi, s)$ son pesos (i.e. *scores*) asociados a las direcciones de la fuente y que dependen del individuo, $H_{av}(f)$ es una función media para cada una de las componentes de frecuencia de la función **H**, a demás representa la media ponderada (i.e. μ) de las HRTFs a lo largo de las M direcciones de las fuentes. [1].

Es importante notar que, $\mathbf{d}_q(f)$ puede ser idéntico o diferente para las HRTFs izquierda y derecha. Por el contrario, $\mathbf{w}_q(\theta, \phi, s)$ siempre difiere para HRTFs izquierdo y derecho. En nuestro proyecto se calcularon $\mathbf{d}_q(f)$ idénticos para ambos oídos. Es por ello que la funciona HRTF solo puede ser definida una vez seleccionada los vectores base $\mathbf{d}_q(f)$ y determinados los pesos $\mathbf{w}_q(\theta, \phi, s)$.

Dado que las HRTFs en cada dirección está representada por N muestras de frecuencias discretas; podemos expresar cada HRTF como: $\mathbf{H}(\theta_i, \phi_i, f_k) = \mathbf{H}(i, k)$ donde i y k representan la dirección y frecuencia discretas, respectivamente. Entonces la Ecuación 1.20, puede ser escrita como:

$$\mathbf{H}(i, k) = \mathbf{w}_q(i) \cdot \mathbf{d}_q(k) + \mathbf{H}_{av}(k) \quad i = 1, 2, ...(M), k = 1, ...(N) \tag{1.21}$$

Los vectores base $\mathbf{d}_q$ mostrados en la Ecuación 1.21, representan los vectores propios de la

matriz de covarianza mencionada en la Sección 1.3.3.2. Para la obtención de las ecuaciones, Ecuación 1.20 y Ecuación 1.21 se empelo el algoritmo descrito en la Sección 1.3.3.2, con las consideraciones ya descritas en esta sección.

Para la reconstrucción de las HRTFs usaremos Q componentes principales, las mismas que están asociadas con los valores propios.Entonces, la representación aproximada de las funciones HRTFs viene dada por la ecuación

$$\mathbf{H_{aprox}}(i,k) = \sum_{q=1}^{Q} \mathbf{w}_q(i) \cdot \mathbf{d}_q(k) + \mathbf{H}_{av}(k) \quad i = 1, ...M, k = 1, ...N \tag{1.22}$$

La Ecuación 1.22 es equivalente a la Ecuación 1.15, explicada en la Sección 1.3.3.3.

2. METODOLOGÍA

En este capítulo se detallarán los procedimientos seguidos para analizar y sintetizar las funciones HRTFs, usando el análisis de componentes principales dentro del entorno de MATLAB. En la Sección 2.1 se presenta la base de datos usada en el proyecto. El procedimiento para la extracción de datos específicos implicados en el desarrollo del proyecto (i.e. creación de tablas en MATLAB) se explica en la Sección 2.1.1. Por otra parte, las etapas de Pre-procesamiento, Análisis y Síntesis presentadas en el alcance, se especifican en las Secciones 2.2, 2.3 y 2.4 respectivamente.

Finalmente se presentará el uso de la validación cruzada (*K-fold cross validation*) como técnica empleada para validar los datos obtenidos en el análisis y síntesis, y así, garantizar que exista independencia entre los datos de entrenamiento y prueba.

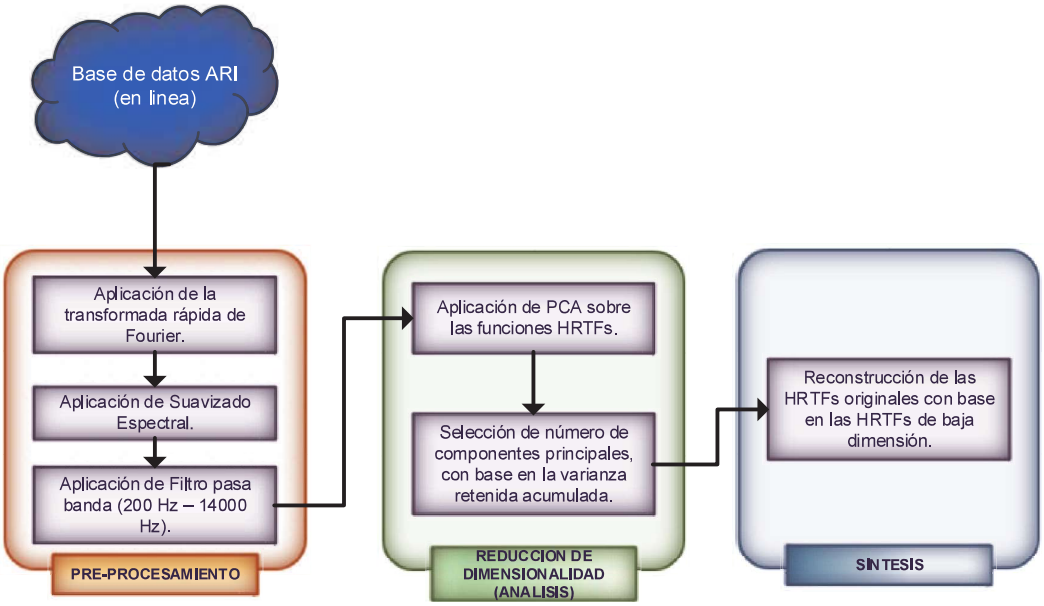

Figura 2.1: Diagrama de bloques.

2.1. BASE DE DATOS ARI

El Instituto de Investigación Acústica (*Acoustic Research Institute*, ARI) es una institución de investigación que forma parte de la Academia de Ciencias de Austria, dedicada a la investigación fundamental en acústica de aplicación abierta [26]. A continuación se presentará una breve descripción de las principales características de la base de datos ARI.

Las HRTFs individuales se midieron en una cámara semi-anecoica. Se usaron 22 altavoces con una respuesta en frecuencia de $\pm$ 4 dB en el rango de 200 a 16000 Hz se montaron en elevaciones fijas de -30° a 80° [26]. Se colocó al sujeto sentado en el centro del arco con micrófonos en el oído (Sennheiser KE-4-211-2) colocados en sus canales auditivos. Se usó

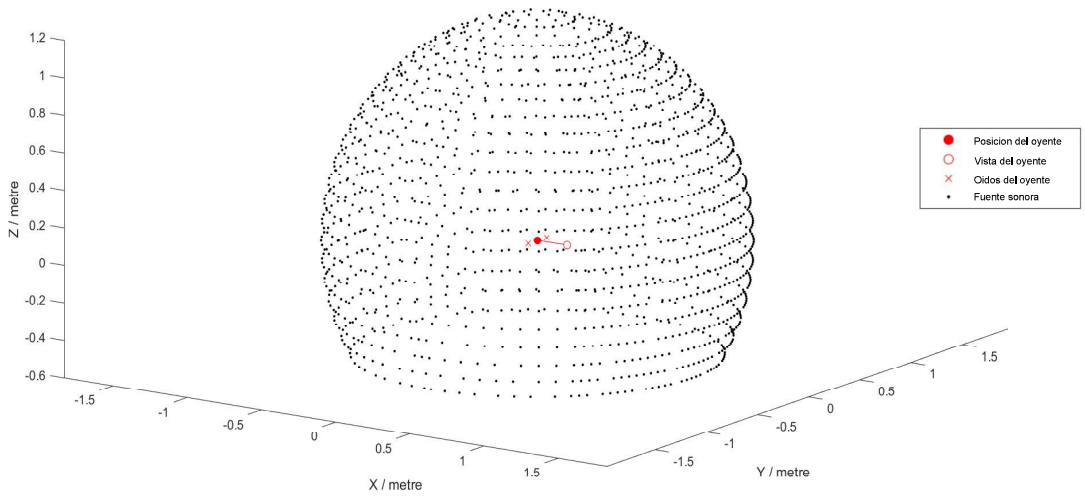

Figura 2.2: Representación de datos obtenidos para un solo sujeto en la base de datos ARI.

un barrido de frecuencia exponencial de 1728.8 ms que comenzó a 50 Hz y finalizó a 20 kHz para medir cada HRTF [26]. Las HRTF se midieron para un azimut y varias elevaciones a la vez, reproduciendo los barridos y grabando las señales en los micrófonos. Luego, el sujeto se giró 2.5° para medir las HRTF para el siguiente azimut.

En el plano horizontal interaural, las HRTF se midieron con un espaciado de 2.5° dentro del rango de azimut de $\pm 45^\circ$ y con un espaciado de 5° fuera de este rango. Las posiciones de las HRTF se distribuyeron con un ángulo esférico constante, lo que significa que el número de HRTF medidas en un plano horizontal disminuyó al aumentar la elevación. Por ejemplo, en la elevación de 80°, solo se midieron 18 HRTF. En total, se midieron 1550 HRTF para cada oyente. Para disminuir el tiempo total requerido para medir las HRTF, se aplicó el método de barrido exponencial múltiple (*Multiple Exponential Sweep Method*, MESM) [17]. Los rangos válidos en la medición se establecieron en 2.5 cm para la posición, 2.5° para el azimut y 5° para la elevación.

2.1.1. EXTRACCIÓN DE DATOS DE LA BASE DE DATOS ARI

Durante el presente proyecto se trabajó con una una base de datos de 95 sujetos con audición normal [35], cuyas HRTFs están respectivamente ecualizadas entre 300 Hz y 18 kHz [35]. El repositorio desde el cual se descargo la base de datos es: `http://sofacoustics.org/data/database/ari/`.

Cabe mencionar que los datos descargados de este repositorio tienen el siguiente formato en su nombre: ***hrtf_nhX.sofa***; donde X representa el ID del sujeto. Esta aclaración se hace debido a que en este repositorio no se encuentran únicamente las funciones HRTFs, sino también, funciones DTF (*Directional Transfer Functions*) que son el resultado de retirar la media de las funciones HRTFs y funciones de transferencia ecualizadas entre 50 Hz y 18

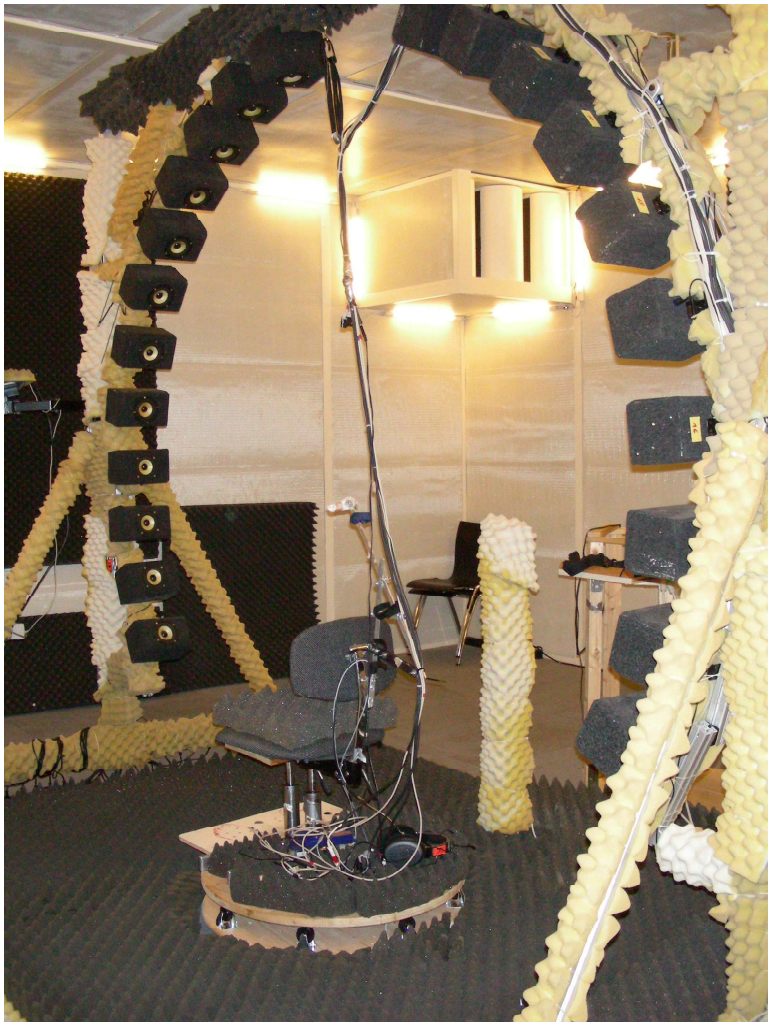

Figura 2.3: Cámara semi-anecoica usada en la obtención de base de datos ARI. Fuente [26].

kHz para auralizaciones de alta fidelidad. En el presente proyecto se usaron solo funciones HRTFs.

Para el manejo de los archivos SOFA, se usó la API (*Application Programming Interface*) *SOFA API for Matlab/Octave* desarrollada por [35], que se encuentra disponible para su descarga de manera gratuita en la plataforma de GitHub. Dentro de esta API encontramos un conjunto de funciones que facilitan el manejo las HRTFs en formato SOFA para MATLAB. Para la extracción de datos fue necesario el seguir el siguiente procedimiento:

1. Descargar el directorio completo de SOFA API for Matlab/Octave en: `https://github.com/sofacoustics/API_MO`.

2. Una vez descargado el directorio, se debe abrir MATLAB y se usará como *Current Folder*, la carpeta que contenga el directorio API_MO, ya que dentro del mismo están las funciones desarrolladas para trabajar en MATLAB con el formato SOFA. En el directorio API_MO se encuentran además: funciones de conversión de coordenadas

(dentro del directorio coordinates), ejemplos (dentro de demo) y test de prueba (dentro de test). Sin embargo, y a pesar de que el API de SOFA presenta una gran variedad de funciones, para este proyecto se usaron esencialmente dos de ellas para la extracción de datos (i.e. SOFAstart.m y SOFAload.m) las cuales se explicarán a continuación.

3. El Segmento de codigo 2.1 muestra el uso de las funciones, SOFAstart.m, la cual inicia el entorno SOFA y añade todas las rutas necesarias y verifica si la versión de MATLAB es adecuada para trabajar con esta API. Esta función debe ser ejecutada antes de cualquier otra al momento de trabajar con archivos en formato SOFA. Una vez iniciado SOFA se procede a cargar las funciones HRTFs, para esto se usó la función SOFAload.m que recibe como parámetro de entrada la dirección completa del archivo .sofa. La función SOFAload entrega un resultado tipo *struct* en el *workspace* de MATLAB, dentro del cual encontraremos todos los datos que maneja el formato SOFA. De especial interés para este proyecto son la variable tipo *struct* de nombre Data, y la variable tipo matriz de nombre SourcePosition.

Segmento de código 2.1: Función SOFAload.m

```
SOFAstart;
HRTF=SOFAload('C:\textbackslash Users \textbackslash ...
    API\_MO-master \textbackslash HRTFs \textbackslash ARI_HRTF ...
    \textbackslash hrtf\_nh.sofa')
```

4. Dentro de la variable Data encontraremos la matriz *IR*, que tiene los valores de respuesta impulsiva medida para un oyente especifico. La matriz *IR* es un arreglo tridimensional de dimensiones $[1550, 2, 256]$, es decir, que para un determinado sujeto, se han tomado 1550 mediciones en distintas ubicaciones de la fuente de sonido para cada oído, y en la que cada medición cuenta con 256 muestras en tiempo discreto.

5. Para un mejor manejo de los datos obtenidos en la matriz IR se optó por separarlos por oído, para lo cual se uso el comando squeeze de MATLAB, de la siguiente manera

Segmento de código 2.2: Uso comando squeeze

```
[squeeze(HRTF.Data.IR(:,1,:))]   %Respuesta impulsiva para el oído ...
    Izquierdo
[squeeze(HRTF.Data.IR(:,2,:))]   %Respuesta impulsiva para el oído ...
    derecho
```

esto permitió separar la matriz IR en dos matrices bidimensionales para los oídos izquierdo y derecho respectivamente.

6. La matriz SourcePosition es una matriz bidimensional con dimensiones [1550,3], que almacena la información de las 1550 posiciones de las fuentes sonoras en coordenadas

esféricas. La primera columna de esta matriz corresponden a los ángulos de azimut, mientras que la segunda columna corresponde a los ángulos de elevación y la tercera columna indica el radio al que se encuentra la fuente de sonido medido desde el centro de la cabeza del oyente.

Para el presente proyecto se extrajeron las dos primeras columnas de la matriz SourcePosition (i.e. las columnas correspondientes a los ángulos azimut y elevación). Ya que como se explicó en la Sección 1.3.2 en este proyecto se usó HRTFs de campo lejano, las mismas que son independientes de la distancia entre le oyente y la fuente de sonido.

2.1.1.1. Construcción de tabla de datos para los 95 sujetos estudiados en el proyecto

El procedimiento descrito en el punto anterior corresponde a la extracción de datos para un solo sujeto, y dado que la base de datos usada cuenta con 95 sujetos, se optó por organizar a los sujetos con sus respectivos datos en una tabla, para lo cual se desarrolló la función extraccion_datos.m, cuyo propósito consiste en extraer los datos de los 95 sujetos, y organizarlos en una tabla. El Segmento de código 2.3 muestra la función desarrollada en MATLAB.

Segmento de código 2.3: Función para extracción y organizacion de datos

```
function tabla = extraccion_datos(path,lista_archivos,numero_muestras)
tamanio=numero_muestras;      %NUMERO DE MUESTRAS POR INDIVIDUO
for j=0:size(lista_archivos,1)-1
    aux=lista_archivos(j+1,:);                %SE LEE LA LINEA 'j' de ...
        lista_archivo
    hrtf=SOFAload(fullfile(path,aux));        %SE CARGA EL ARCHIVO *.sofa
    %SE EXTREA LA MATRIZ DE IR
    HRIR(j*tamanio+1:(j+1)*tamanio,:)=[squeeze(hrtf.Data.IR(:,1,:)); ...
    squeeze(hrtf.Data.IR(:,2,:))];
    %SE EXTREA LAMMATRIZ DE AZIMUT
    AZIMUT(j*tamanio+1:(j+1)*tamanio,:)=[hrtf.SourcePosition(:,1); ...
    hrtf.SourcePosition(:,1)];
    %SE EXTREA LAMMATRIZ DE ELEVACION
    ELEVACION(j*tamanio+1:(j+1)*tamanio,:)=[hrtf.SourcePosition(:,2); ...
    hrtf.SourcePosition(:,2)];
    %SE CREA LA MATRIZ DE OIDOS
    EAR(j*tamanio+1:(j+1)*tamanio,:)=[ones(tamanio/2,1);2*ones(tamanio/2,1)];
    %SE AGRAGA EL ID DEL SUJETO
    SUJETO(j*tamanio+1:(j+1)*tamanio,:)=(j+1)*ones(tamanio,1);
end
 tabla=table(SUJETO,AZIMUT,ELEVACION,EAR,HRIR);
end
```

El resultado entregado por la función extraccion_datos.m es una tabla con las siguientes características:

- Cuenta con cinco columnas, SUJETO, AZIMUT, ELEVACION, EAR, HRIR, que corresponden al ID del sujeto, el ángulo de azimut, el ángulo de elevación, el oído correspondiente(se usaron los números 1 para oído izquierdo y 2 para oído derecho) y la respuesta impulsiva (IR) de cada sujeto, respectivamente.

Tabla 2.1: Ejemplo de tabla para un solo sujeto.

SUJETO	AZIMUT	ELEVACION	EAR	HRIR
$[ID^{(1)}]_{1550x1}$	$[AZ^{(1)}]_{1550x1}$	$[ELV^{(1)}]_{1550x1}$	$[1]_{1550x1}$	$[IR_I^{(1)}]_{1550x256}$
$[ID^{(1)}]_{1550x1}$	$[AZ^{(1)}]_{1550x1}$	$[ELV^{(1)}]_{1550x1}$	$[2]_{1550x1}$	$[IR_D^{(1)}]_{1550x256}$

- Cada sujeto cuenta con 3100 posiciones de azimut y elevación (que corresponden a un número similar de filas dentro de la tabla), esto debido a que primero se presentan las 1550 respuestas impulsivas para el oído izquierdo, y a continuación las 1550 respuestas impulsivas para el oído derecho (ver Tabla 2.1).

- Los datos correspondientes a cada sujeto han sido colocados uno debajo de otro (ver Tabla 2.2), haciendo que en total la tabla cuente con 294500 filas.

Tabla 2.2: Formato de tabla obtenida con el script al extraer datos de la base de datos ARI.

SUJETO	AZIMUT	ELEVACION	EAR	HRIR
$[ID^{(1)}]_{3100x1}$	$[AZ^{(1)}]_{3100x1}$	$[ELV^{(1)}]_{3100x1}$	$[EAR^{(1)}]_{3100x1}$	$[IR^{(1)}]_{3100x256}$
$[ID^{(2)}]_{3100x1}$	$[AZ^{(2)}]_{3100x1}$	$[ELV^{(2)}]_{3100x1}$	$[EAR^{(2)}]_{3100x1}$	$[IR^{(2)}]_{3100x256}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[ID^{(94)}]_{3100x1}$	$[AZ^{(94)}]_{3100x1}$	$[ELV^{(94)}]_{3100x1}$	$[EAR^{(94)}]_{3100x1}$	$[IR^{(94)}]_{3100x256}$
$[ID^{(95)}]_{3100x1}$	$[AZ^{(95)}]_{3100x1}$	$[ELV^{(95)}]_{3100x1}$	$[EAR^{(95)}]_{3100x1}$	$[IR^{(95)}]_{3100x256}$

Una vez extraídos y almacenados los datos en la Tabla 2.2, se procedió con las siguientes etapas del proyecto.

2.2. PRE-PROCESAMIENTO DE DATOS

Dentro de la etapa de pre-procesamiento de datos se deben realizar tres procesos. Primero se aplicará la transformada rápida de Fourier sobre los datos, a continuación se procederá a aplicar suavizado espectral, y filtrado espectral. Estos procesos se describen a continuación.

2.2.1. APLICACIÓN DE LA TRANSFORMADA RÁPIDA DE FOURIER

Se aplicó la transformada rápida de Fourier a la columna HRIR de la Tabla 2.2 a fin de obtener las funciones HRTFs. Para lo cual se usó la función fft de MATLAB. A continuación se presenta el código empleado, para el cual se tomaron las siguientes consideraciones:

Segmento de código 2.4: Aplicacion de la transformada de Fourier en MATLAB

```
%%APLICACION DE LA TRANSFORMADA RAPIDA DE FOURIER
HRIR=TABLA_NUEVA.HRIR;
%Se aplica la funcion fft a lo largo de cada fila
HRTF_x_two_slided=abs(fft(HRIR,[],2));
%Se toma unicamente la mitad de los datos de la fft
HRTF_x_single_slided=HRTF_x_two_slided(:,1:floor(N/2)+1);
%Se multiplica todos los valores por dos excepto la componente DC
HRTF_x_single_slided(:,2:end-1)=2*HRTF_x_single_slided(:,2:end-1);
%Se agrega a la tabla la colunma HRTF
TABLA_NUEVA.HRTF=HRTF_x_single_slided;
TABLA_HRTF=TABLA_NUEVA;
save('DATOS_HRTF.mat','TABLA_HRTF')
```

- La transformada rápida de Fourier se aplicó a lo largo de cada fila.
- Durante el proyecto se trabajo con funciones de fase mínima ya que como se describió en la Sección 1.3.2.5 esta aproximación presenta una gran ventaja al permitir procesar las HRTFs utilizando únicamente su magnitud. Por este motivo, se extrajo el modulo de las funciones HRTFs con el comando *abs* de MATLAB.
- Una vez obtenida la transformada rápida de Fourier, se agregó la misma como una columna adicional (de nombre HRTF) a la Tabla 2.2 como muestra la Tabla2.3.

Tabla 2.3: Tabla de Datos incluida la HRTF

SUJETO	AZIMUT	ELEVACION	EAR	HRIR	HRTF
$[ID^{(1)}]_{3100x1}$	$[AZ^{(1)}]_{3100x1}$	$[ELV^{(1)}]_{3100x1}$	$[EAR^{(1)}]_{3100x1}$	$[IR^{(1)}]_{3100x256}$	$[HRTF^{(1)}]_{3100x129}$
$[ID^{(2)}]_{3100x1}$	$[AZ^{(2)}]_{3100x1}$	$[ELV^{(2)}]_{3100x1}$	$[EAR^{(2)}]_{3100x1}$	$[IR^{(2)}]_{3100x256}$	$[HRTF^{(2)}]_{3100x129}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[ID^{(94)}]_{3100x1}$	$[AZ^{(94)}]_{3100x1}$	$[ELV^{(94)}]_{3100x1}$	$[EAR^{(94)}]_{3100x1}$	$[IR^{(94)}]_{3100x256}$	$[HRTF^{(94)}]_{3100x129}$
$[ID^{(95)}]_{3100x1}$	$[AZ^{(95)}]_{3100x1}$	$[ELV^{(95)}]_{3100x1}$	$[EAR^{(95)}]_{3100x1}$	$[IR^{(95)}]_{3100x256}$	$[HRTF^{(95)}]_{3100x129}$

2.2.2. APLICACIÓN DE FILTRADO Y SUAVIZADO ESPECTRAL

Una vez obtenida la transformada rápida de Fourier, se procedió a aplicar suavizado espectral en la columna **HRTF**. Para lo cual se usó la función $\mathrm{HRTFsmoothing.m}$. El suavizado espectral en las HRTFs se aplicó con anchos de banda de 1 ERB (*Equivalent Rectangular Bandwidth*) para frecuencias bajas (i.e. f <5 kHz) y 2 ERB para y altas frecuencias (i.e. f $\geq$ 5 kHz). Estos parámetros de suavizado fueron escogidos de acuerdo a los resultados de Xie et al [1] que demuestran que es posible suavizar las HRTFs con los valores de ERB mencionados sin que el oyente pierda precisión en su percepción auditiva.

Posteriormente y sobre la misma columna se aplicó un filtro ideal pasa banda entre las frecuencias de 200 Hz y 14000 Hz, usando la función $\mathrm{filtrado_HRTF.m}$, esto con el fin de reducir

los efectos producidos por las limitaciones en la respuesta de frecuencia de los equipos utilizados para la medición de HRIRs [34]. Con la aplicación de este filtro se pasó de tener 129 a 75 componentes en frecuencia.

A continuación se presenta el segmento de código correspondiente al suavizado y filtrado espectral.

Segmento de código 2.5: Suavizado Espectral y filtrado pasa banda

```
%%SUAVIZADO ESPECTRAL Y FILTRADO PASA BANDA ENTRE 200 Hz y 14000Hz
DATOS=struct('fmin',200,'fmax',14000,'Fs',48000);
%Funcion de suavizado espectral
HRTF_SUAVIZADO=HRTFsmoothing(TABLA_HRTF.HRTF,48000);
%Funcion de filtrado espectral
HRTF_FILTRADO=filtrado_HRTF(HRTF_SUAVIZADO,DATOS);
%Se agrega la columna de HRTF_FILTRADO_SUAVIZADO
TABLA_HRTF.HRTF_FILTRADO_SUAVIZADO=HRTF_FILTRADO;
%Armado de nueva tabla de datos
BASE_DATOS=TABLA_HRTF;
BASE_DATOS.HRIR=[];
BASE_DATOS.HRTF=[];
%Se guarda la nueva tabla
save('DATOS_PROCESADOS.mat','BASE_DATOS')
```

Las funciones HRTFsmoothing.m y filtrado_HRTF.m, son presentadas a continuación. Es importante mencionar que la función HRTFsmoothing fue obtenida de [36].

Segmento de código 2.6: Función de filtrado espectral filtrado_HRTF.m

```
function HRTF_reducido = filtrado_HRTF(X,PARAMETROS)
%EN EL CAMPO "PARAMETROS" SE DEBEN ESPECIFICAR LOS SIGUEINTES TRES DATOS
%fmin: FRECUENCIA DE CORTE MINIMA DEL FILTRO
%fmax: FRECUENCIA DE CORTE MAXIMA DEL FILTRO
%Fs: FRECUENCIA DE MUESTREO
lim_min=floor(PARAMETROS.fmin/(PARAMETROS.Fs/2)*size(X,2))+1;
lim_max=floor(PARAMETROS.fmax/(PARAMETROS.Fs/2)*size(X,2)+1);
HRTF_reducido=X(:,lim_min:lim_max);
end
```

Segmento de código 2.7: Función de suavizado espectral HRTFsmoothing.m

```
function Y = HRTFsmoothing(X, fs)
%%this code expects all band from 0 to fs/2.ERB filtering
%X is the magnitude (not log magnitude), columns are coefi, row
%observations
%%opts.fs should be freq sampling.

```

```
%%
L = size(X,2); %filter length
%N = size(X,1); %num of obs
Xsquared = X.^2;

fmin = 0;
fmax = fs/2;
df = (fmax-fmin)/L;
f = linspace(fmin, fmax, L);
ERB = 24.7*(4.37*f/1000+1);
ERBN = 21.4*log10(4.37*f/1000 + 1);
%fl = f-ERB/2
%fh = f+ERB/2
%%
Ysquared = zeros(size(X));
fc = 5000; %Cut frequency for high freq
W = round(1*ERB/df); %by default we use 1 ERB.
W = round(1*ERB/df).*(f<fc) + round(2*ERB/df).*(f≥fc); %ERB for high and low ...
    freq is different, HF we use 2 ERB
W(W==0)=1; %Window cannot be zero
W(¬mod(W,2))=W(¬mod(W,2))-1;%Window must be odd
for j=1:L
    if j≤floor(W(j)/2) %left border
            temp = mean(Xsquared(:,1:j+floor(W(j)/2)),2);
        else
            if (j+floor(W(j)/2))>L %right border
                temp = mean(Xsquared(:,j-floor(W(j)/2):end),2);
            else %in the middle
                temp = mean(Xsquared(:,j-floor(W(j)/2):j+floor(W(j)/2)),2);
            end
    end
    Ysquared(:,j)= temp;
end
Y = sqrt(Ysquared);
```

Una vez aplicado el filtrado y suavizado espectral se procedió a retirar las columnas **HRIR** y **HRTF** de la Tabla 2.3 y se ha añadio la columna **HRTF_FITRADO_SUAVIZADO** como muestra la Tabla 2.4. Esta última, contiene las HRTFs obtenidas al finalizar la etapa de pre-procesamiento.

Tabla 2.4: Tabla de Datos incluida la HRTF

SUJETO	AZIMUT	ELEVACION	EAR	HRTF_FITRADO_SUAVIZADO
$[ID^{(1)}]_{3100x1}$	$[AZ^{(1)}]_{3100x1}$	$[ELV^{(1)}]_{3100x1}$	$[EAR^{(1)}]_{3100x1}$	$[HRTF^{(1)}]_{3100x75}$
$[ID^{(2)}]_{3100x1}$	$[AZ^{(2)}]_{3100x1}$	$[ELV^{(2)}]_{3100x1}$	$[EAR^{(2)}]_{3100x1}$	$[HRTF^{(2)}]_{3100x75}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[ID^{(94)}]_{3100x1}$	$[AZ^{(94)}]_{3100x1}$	$[ELV^{(94)}]_{3100x1}$	$[EAR^{(94)}]_{3100x1}$	$[HRTF^{(94)}]_{3100x75}$
$[ID^{(95)}]_{3100x1}$	$[AZ^{(95)}]_{3100x1}$	$[ELV^{(95)}]_{3100x1}$	$[EAR^{(95)}]_{3100x1}$	$[HRTF^{(95)}]_{3100x75}$

La Figura 2.4 presenta una comparación entre la HRTF original con la HRTF obtenida luego del filtrado y suavizado espectral. En esta podemos observar que el suavizado espectral tiene mayor efecto en frecuencias superiores a los 4 kHz.

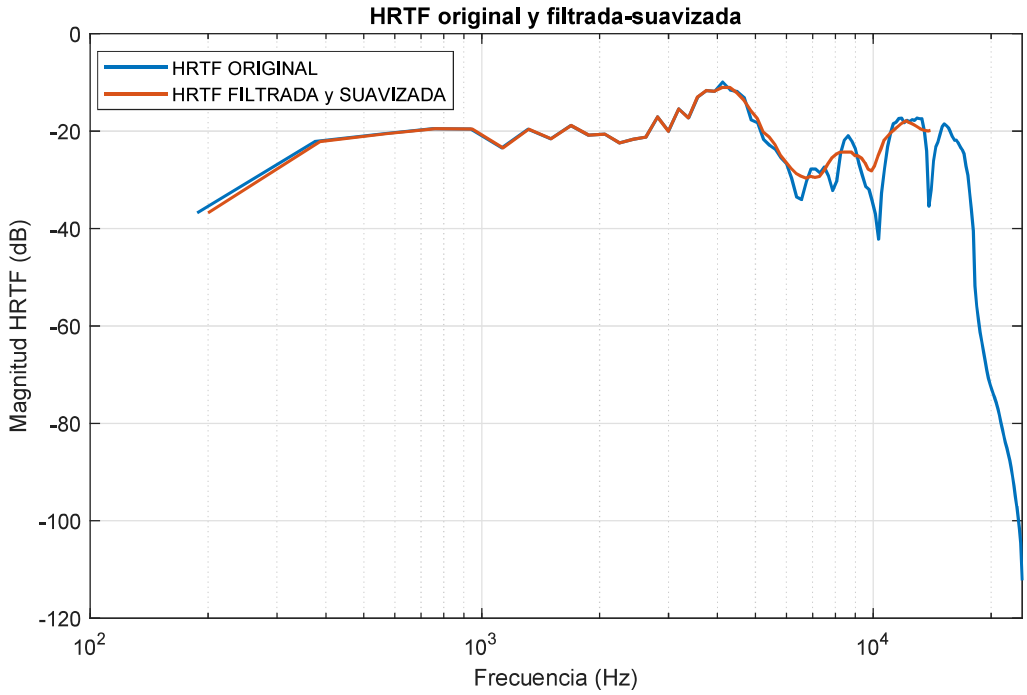

Figura 2.4: Comparación entre las HRTF originales y HRTF suavizada y filtrada para un mismo sujeto y una misma dirección.

2.3. ANÁLISIS DE HRTFs USANDO PCA

Dentro de esta etapa se analizará como varían las componentes principales con respecto a la ubicación de la fuente de sonido. Para esto es necesario aplicar PCA en la columna **HRTF_FITRADO_SUAVIZADO** y escoger un número adecuado de componentes principales que revelen características importantes de las funciones HRTFs.

2.3.1. USO DE PCA EN MATLAB

Se usó la función pca de MATLAB, la misma que se encarga del Análisis de Componentes Principales de datos no procesados, esta función acepta como argumento de entrada una matriz **X** de dimensiones $m \times n$, cuyas m filas corresponderán a observaciones, y sus n columnas serán sus respectivas variables (i.e. n componentes en frecuencia de las HRTFs) , además, la función pca es capaz de entregar como resultado los siguientes argumentos de salida:

Segmento de código 2.8: Argumentos de entrada y salida PCA

```
[coeff,score,latent,tsquared,explained,mu] = pca(X)
```

- **coeff:** La matriz coeff es una matriz cuadrada de tamaño $n \times n$, en la cual, cada columna de la matriz contiene los coeficientes para un componente principal. Las columnas están ordenadas en forma descendente en función de su varianza retenida. De forma predeterminada, pca centra los datos (i.e. normaliza la media a cero) y utiliza el algoritmo de descomposición de valores singulares (SVD).
- **score:** Es una matriz de dimensiones iguales a **X** (i.e. $m \times n$), que representan la proyección de los datos originales en el nuevo espacio vectorial.
- **latent:** Entrega los valores propios de la matriz de covarianza de **X**, devueltos como un vector columna de tamaño $n \times 1$ que representan los valores de varianza retenida para cada componente principal.
- **tsquared:** Presenta la suma de los cuadrados de los score estandarizadas para cada observación, se devuelve como un vector columna de tamaño $m \times 1$.
- **explained:** Representa el porcentaje de la varianza total explicada por cada componente principal, devuelto como un vector columna e tamaño $n \times 1$.
- **mu:** Representa el valor de la media de cada componente principal,entregado como un vector fila tamaño $1 \times n$, donde el i-ésimo elemento de mu, es la media aritmética de la i-ésima variable de la matriz **X**.

Para la etapa de análisis de este proyecto se usaron únicamente cuatro de los argumentos de salida de la función pca, como se muestra en el Segmento de codigo 2.9.

Segmento de código 2.9: Aplicación de PCA sobre las HRFTs

```
%%APLICANDO PCA SOBRE LA BASE DE DATOS
[coeff,score,varianzas,¬,¬,mu]=pca(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO);
```

donde, $\mathrm{BASE_DATOS.HRTF_FILTRADO_SUAVIZADO}$, es una matriz que contiene las HRTFs originales de $m = 294500$ filas correspondientes a las observaciones de los 95 sujetos estudiados y $n = 75$ columnas que corresponden a las variables.

Una vez ejecutado el Segmento de Código 2.9 se obtuvieron cuatro matrices, coeff, score,$\mathrm{varianzas}$ y mu con las carácter descritas en la Tabla 2.5

2.3.2. SELECCIÓN DE NÚMERO DE COMPONENTES PRINCIPALES

Para seleccionar el número de componentes principales que se van a usar durante el resto del proyecto, se usó la varianza retenida presente en la matriz $\mathrm{varianzas}$, la misma que sirvió como base para obtener la varianza retenida acumulada, obtenida con la Ecuación 1.19,

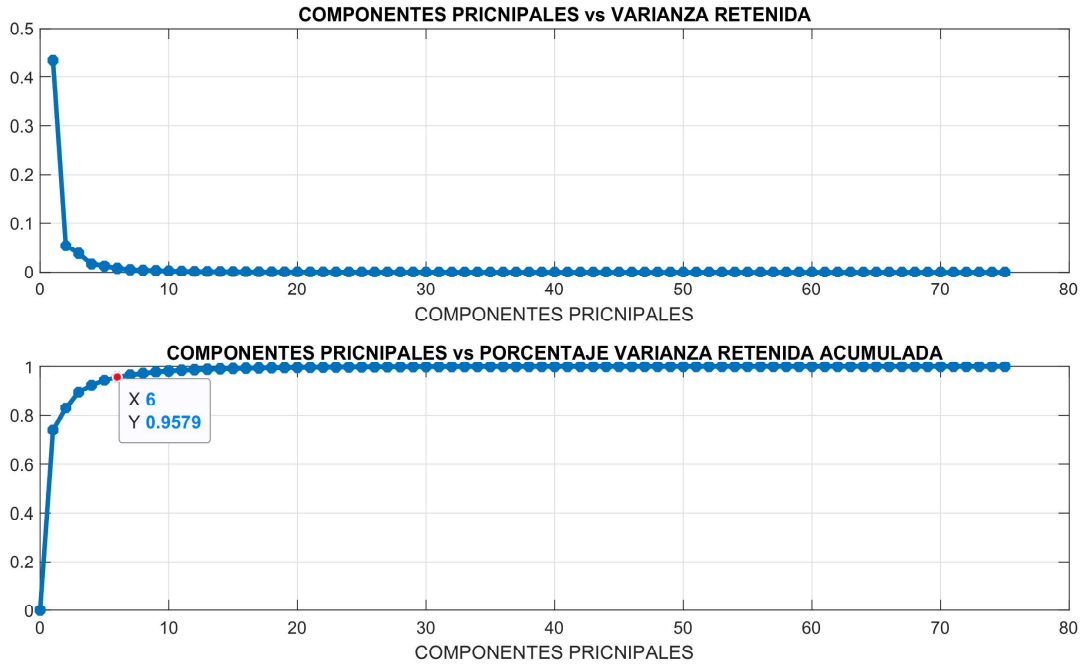

Figura 2.5: Análisis de componentes principales vs la varianza retenida acumulada.

cuya función podemos ver graficada en la Figura 2.5.b. En contraste, la Figura 2.5.a, muestra la varianza retenida individual de cada componente principal. De esta última es importante destacar como las primeras componentes principales han conservado la mayor cantidad de características de las HRTFs originales. Para seleccionar el número de componentes principales se buscó conservar al menos el 95 % de la información de las HRTFs originales con el menor número de componentes principales posible, a fin de maximizar la reducción de dimensionalidad.

La Tabla 2.6 muestra el porcentaje de varianza retenida acumulada correspondientes a las 10 primeras componentes principales. En la misma podemos ver como con 6 componen-

Tabla 2.5: Descripción de matrices obtenidas con pca

Matriz	Descripción	Dimensiones
coeff	Matriz cuadrada, en la que cada columna corresponde a una de las 75 componentes principales de BASE_DATOS.HRTF_FILTRADO_SUAVIZADO.	75x75
score	Matriz que contiene los valores proyectados en el nuevo espacio vectorial de la matriz, BASE_DATOS.HRTF_FILTRADO_SUAVIZADO	294500x75
varianzas	Vector columna, que contiene las varianzas retenidas de cada componente principal	75x1
mu	Vector fila, con la media aritmética correspondiente a cada columna.	1x75

Tabla 2.6: Porcentaje de Varianza Retenida Acumulada

Número de Componentes Principal	Porcentaje de Varianza Retenida Acumulada
1	0.7396
2	0.8305
3	0.89.52
4	0.9236
5	0.9448
6	0.9579
7	0.9663
8	0.9728
9	0.9783
10	0.9821

tes principales conseguimos el 95.79 % de varianza retenida, alcanzando así el objetivo de retener al menos el 95 % de la información de las HRTFs originales.

Con esto hemos logrado reducir la dimensionalidad de los datos originales, reduciendo las 75 variables originales a 6 componentes principales. A partir de ahora y tanto para los experimentos de síntesis y análisis se usarán estas seis componentes principales. En el Capitulo 3 analizaremos como las componentes principales varían con respecto a la posición de la fuente de sonido. Finalmente es importante resaltar que los valores obtenidos en la matriz score han sido normalizados con medioa cero y varianza uno con la función zscore de MATLAB.

2.4. SÍNTESIS DE HRTFs CON PCA

2.4.1. RECONSTRUCCIÓN DE LAS HRTFs ORIGINALES CON BASE EN LAS HRTFs DE BAJA DIMENSIÓN.

Una vez que se ha seleccionado el número de componentes principales, con los cuales se va a trabajar durante el resto del proyecto, se procedió a realizar la síntesis de funciones HRTFs. Para lo cual se hizo uso de la Ecuación 1.22, la misma que describe la reconstrucción aproximada de las HRTFs a partir de las componentes principales seleccionadas en la etapa de análisis. Otra forma de ver la síntesis de HRTFs con PCA es con base en el diagrama de bloques mostrado en la Figure 2.6. En este diagrama observamos que la síntesis (i.e. reconstrucción) de las funciones HRTFs requiere de tres sub-etapas esencialmente.

Para la síntesis se hizo uso de las matrices score, coeff y mu, que tienen dimensiones 294500×75 , 75×75 y 1×75 respectivamente. Sin embargo y dado que se usaron únicamente seis componentes principales las matrices score y coeff redujeron sus dimensiones a 294500×6 y 75×6 respectivamente.

A continuación se explicará a detalle el diagrama de bloques de la Figura 2.6.

- El primer bloque de la Figura 2.6 (i.e. Multiplicación Matricial), muestra una multiplicación matricial entre las matrices score y coeff, es importante destacar que previa a esta multifunción, la matriz coeff fue transpuesta. Esto debido a que como se mencionó en la Tabla 2.5, las componentes principales vienen dadas como columnas en la matriz coeff. El resultado de esta multiplicación matricial es una matriz de dimensiones 294500×75, que denominaremos $\mathbf{H}'$.

- Si bien las dimensiones de la matriz $\mathbf{H}'$ son iguales a las de la matriz original (i.e. $\mathrm{BASE_DATOS.HRTF_FILTRADO_SUAVIZADO}$). Esta matriz no es la representación aproximada de las HRTFs originales. Es por esto que en el segundo bloque de la Figura 2.6 (i.e. Suma de media aritmética para cada variable) se suma la media aritmética que fue extraída de cada variable en la etapa de pre-procesamiento. Para ello se sumó cada fila de la matriz $\mathbf{H}'$ con el vector mu obtenido en la etapa de análisis.

- Finalmente en el tercer bloque de la Figura 2.6 (i.e. HRTFs aproximadas) se obtuvo la representación aproximada de las HRTFs originales en la matriz $\mathrm{HRTFapprox}$ la cual cuenta con las mismas 294500 muestras y las 75 variables originales .

Ahora bien, dado que, el propósito final de la síntesis es reconstruir las funciones HRTFs de manera fidedigna, es necesario usar la distorsión espectral como métrica para determinar el error que existe entre la reconstrucción aproximada de las funciones HRTFs y las HRTFs originales.

Podemos calcular la distorsión espectral de la siguiente manera

$$\mathbf{SD}(f) = \left| 20 \log_{10} \frac{|\mathbf{H}|}{|\mathbf{H}_{aprox}|} \right| (dB) \tag{2.1}$$

donde, $\mathbf{H}$ representa las funciones HRTFs originales y $\mathbf{H}_{aprox}$ las HRTFs aproximadas. La distinción espectral especifica la desviación entre las magnitudes de las HRTFs originales

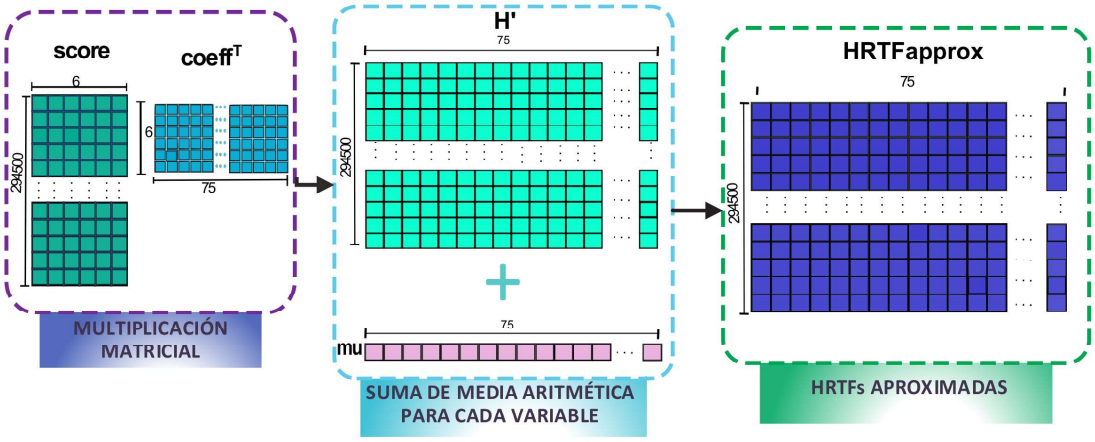

Figura 2.6: Diagrama de bloques para la sintesis de funciones HRTFs usando PCA.

y aproximadas en una escala logarítmica. Una distorsión espectral cercana a 0 dB significa menos desviación [1]. De igual manera es posible definir la distinción espectral media como

$$\mathbf{SD_M}(f) = \sqrt{\frac{1}{N_f}\sum_{k=1}^{N_f}\left(20\log_{10}\frac{|\mathbf{H}|}{|\mathbf{H}_{aprox}|}\right)^2}(dB) \tag{2.2}$$

donde, N_f es igual al número de puntos frecuencia de **H** (i.e. las frecuencias de las funciones HRTFs).

Si bien dentro de este proyecto se seleccionó seis componentes principales para la reconstrucción de las funciones HRTFs, también se realizo la reconstrucción de las mismas con una, dos, tres, cuatro y cinco componentes principales (siguiendo el mismo procedimiento descrito en esta sección). Esto con el fin de determinar como varia la distorsión espectral en función del número de componentes principales.

2.4.2. K-FOLD CROSS-VALIDATION

Una vez realizada la síntesis es necesario evaluar cuan eficiente resulta aplicar PCA sobre las funciones HRTFs. Esto con el fin de validar los resultados obtenidos de este proyecto. Para evaluar la eficiencia de un método cualquiera se puede dividir el conjunto de datos originales en conjuntos de entrenamiento y pruebas. Luego se utiliza los conjuntos de entrenamiento para entrenar el modelo y el conjunto de test para probar el modelo. Luego, evaluamos el rendimiento del modelo en función de una métrica de error para determinar la precisión del mismo.

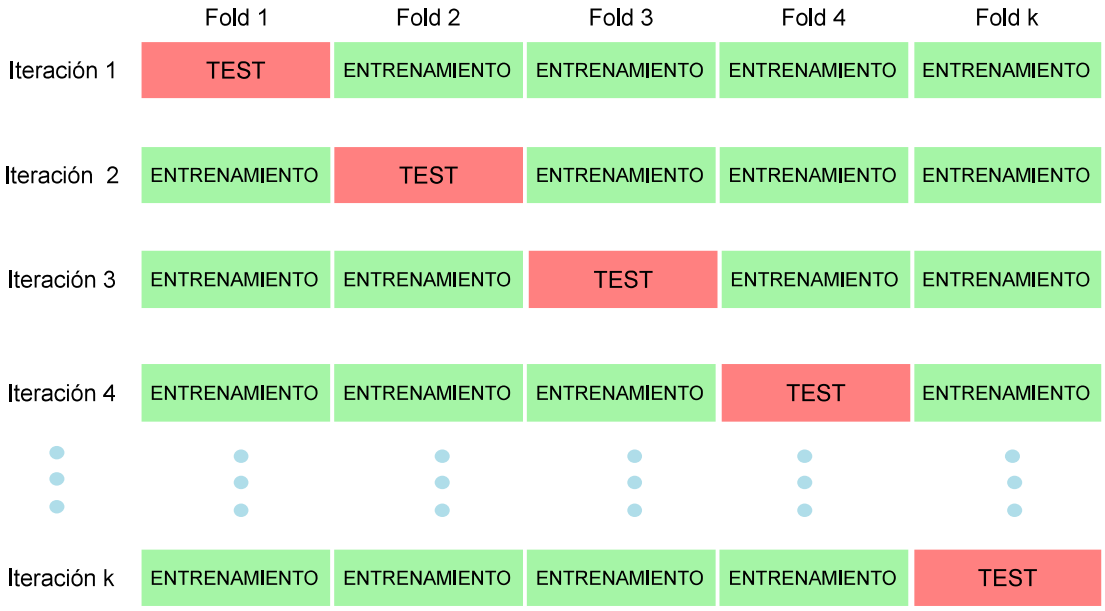

Figura 2.7: División de datos en validación cruzada de k-fold.

Sin embargo, este método no resulta ser del todo confiable, ya que, la métrica de error obtenida para un conjunto de test puede ser diferente a la precisión obtenida para un conjunto

de test diferente. La validación cruzada de k folds (*k-fold cross-validation*) proporciona una solución a este problema ya que divide los datos originales en k-*fold* o k-grupos y garantizar que cada *fold* se utilice como un conjunto de test en algún momento como muestras la Figura 2.7. Es por esto que se usó *k-fold cross-validation*, como técnica para evaluar el rendimiento del modelo PCA en función de una métrica de error (i.e. distorsión espectral) y así determinar la precisión del modelo empleado.

Para aplicar *k-fold cross-validation* dentro del presente proyecto se usaron $k = 5$ folds, entonces, los 95 sujetos de la base de datos se dividieron de forma aleatoria en grupos de 19 sujetos, cada uno correspondiente a un fold. Posteriormente se procedió a aplicar el algoritmo de PCA con cada fold de entrenamiento y se lo almacenó en una tabla junto con sus correspondientes ángulos de azimut y elevación, así como con los indicadores que corresponden a cada oído. El procedimiento descrito puede verse en el segmento de código 2.10.

Segmento de código 2.10: Aplicación k-fold cross-validation

```
%APLICACION DE K-FOLD CROSS VALIDATION
fold=5;          %Número de Grupos en los que se divide los datos originales
[sujetos,indice]=indices(DatosValidacion.SUJETO,fold,3100); %Funcion ...
    indices devuelve las matrices sujetos  e indices, donde sujeto es una ...
    matriz que cuenta con los sujetos y a que indice pertenecen, indice solo ...
    muestra los indices
longitud_test=length(DatosValidacion.SUJETO)/fold; %Numero de sujetos de en ...
    fold
numero_componentes=6; % Componentes principales
for i = 1:fold
    test = (indice == i);
    train = ¬test;
    mat_test((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.HRTF_FILTRADA_SUAVIZADA(test,:);
    mat_azimut((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.AZIMUT(test,:);
    mat_elevacion((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.ELEVACION(test,:);
    mat_ear((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.EAR(test,:);
    [MATRIZ_ROTACIONAL,¬,¬,¬,¬,MEDIA]=...
    pca(DatosValidacion.HRTF_FILTRADA_SUAVIZADA(train,:));
    PROYECCIONES=(DatosValidacion.HRTF_FILTRADA_SUAVIZADA(test,:)-MEDIA)...
        *MATRIZ_ROTACIONAL(:,1:numero_componentes);
    RECONSTRUCCION=PROYECCIONES*MATRIZ_ROTACIONAL(:,1:numero_componentes)'+ ...
        MEDIA;
    mat_recon((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=RECONSTRUCCION;
    FOLD((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
    i*ones(longitud_test,1);
end
```

Tras haber ejecutado el Segmento de código 2.10 se obtuvo una tabla con las siguientes

columnas, SUJETO, AZIMUT, ELEVACION, EAR, FOLD, HRTF_ORIGINAL, HRTF_APROX y SD, las mismas que corresponden respectivamente al identificador del sujeto, los ángulos de elevación y azimut correspondientes a la fuente de sonido para cada sujeto, el indicador del oído, el grupo o *fold* al que perteneces (que pueden ser números del uno al cinco), así como las HRTFs originales y aproximadas. Finalmente se incluye una columna con la distorsión espectral correspondiente a cada observación.

3. RESULTADOS Y DISCUSIÓN (APLICACIÓN METODOLÓGICA)

En este capitulo se presentan los resultados obtenidos de las etapas de análisis y síntesis para las funciones HRTFs reconstruidas con PCA, usando como métrica de error la distorsión espectral estimada usando *k-fold cross validation*.

3.1. RESULTADOS OBTENIDOS EN LA ETAPA DE ANÁLISIS

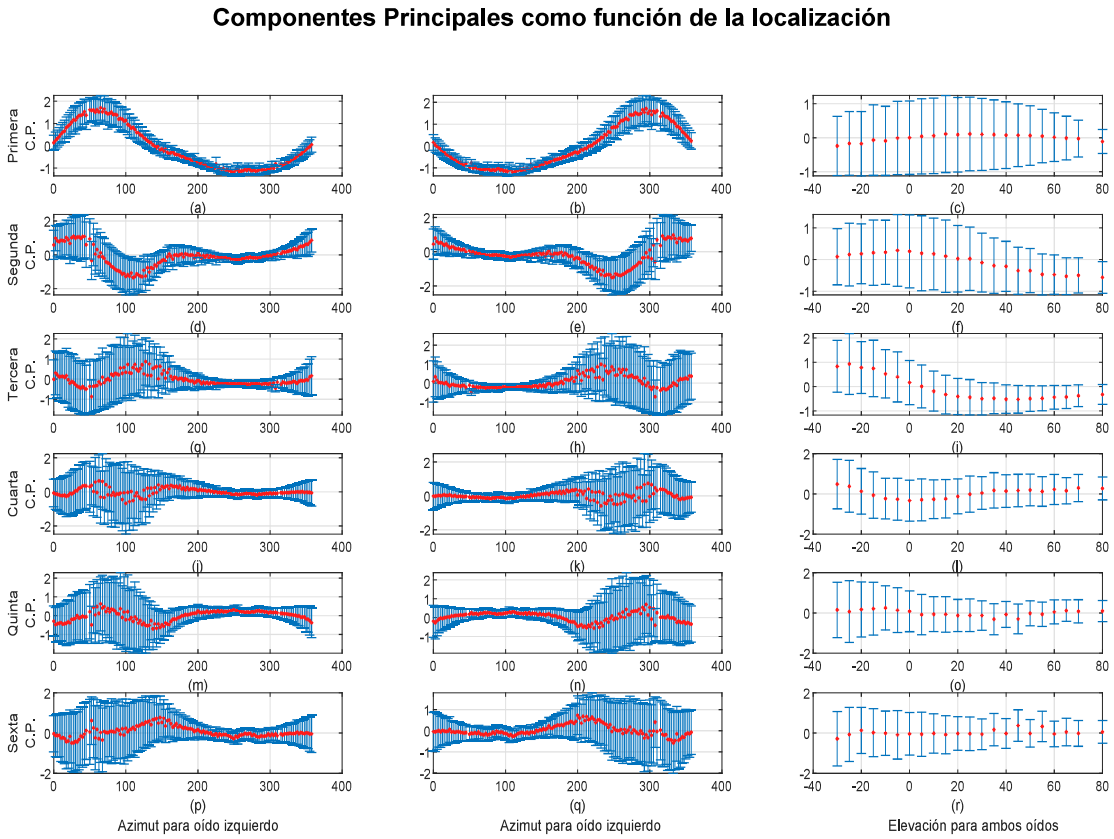

Figura 3.1: Componentes principales en función de la ubicación.

La Figura 3.1, presenta las componentes principales en función de la ubicación para los 95 sujetos estudiados dentro de este proyecto. En todos los gráficos de la figura, los puntos de datos (i.e. puntos rojos) representan la media de todos los sujetos para una ubicación específica y las barras de error corresponden a un intervalo de $\pm$ una desviación estándar .

Los datos de azimut fueron separados en gráficos de oído izquierdo y derecho, esto con el fin de evidenciar la variabilidad que existe entre el ipsilateral y contralateral de los oyentes.

Las Figuras 3.1.a, 3.1.b y 3.1.c, corresponden a la primera componente principal (i.e. la de mayor varianza retenida). En la Figura 3.1.a, que analiza los ángulos de azimut para el oído izquierdo, se observa que existe una alta correlación entre la primera componente principal y

Tabla 3.1: Coeficientes de correlación de las componentes principales

Número de Componentes Principales	Coeficiente de correlación para los ángulos de azimut del oído izquierdo	Coeficiente de correlación para los ángulos de azimut del oído derecho	Coeficiente de correlación para los ángulos de elevación (ambos oídos)
1	-0.7307	0.7148	0.0671
2	-0.0769	0.0016	-0.2287
3	-0.0653	0.0895	-0.4673
4	-0.0087	0.0190	0.0393
5	0.0928	-0.0538	-0.0829
6	0.0119	-0.0281	0.0581

los ángulos de azimut. Esto se hace evidente si comprobamos el coeficiente de correlación que para este caso es $-0{,}7307$. Podemos ver como en esta figura la primera componte principal describe con valores positivos los ángulos pertenecientes al ipsilateral del oyente (i.e. propiamente los ángulos comprendidos entre 0° y 137.5 ° son positivos), mientras que los ángulos correspondientes al contralateral son valores negativos. En la Figura 3.1.b se observa un caso similar, con la diferencia que esta figura analiza los ángulos de azimut para el oído derecho, por lo tanto, los ángulos correspondientes al ipsilateral en este caso son los ángulos mas cercanos al lado derecho de la cabeza del sujeto (i.e. ángulos superiores a 180°), que son descritos con valores positivos . Así mismo, es apreciable un alto grado de correlación entre la primera componente principal y los ángulos de azimut en el oído derecho que en este caso es igual a $0{,}7148$. Por otra parte, en la Figura 3.1.c no se encuentra un patrón especifico para los ángulos de elevación, presentando más bien una respuesta plana para los mismos. Los patrones descritos por las tres figuras sugieren que la primera componente principal puede ayudar a describir las ubicaciones izquierda y derecha, sin embargo entrega poca información de los ángulos de elevación.

Las Figuras 3.1.d y 3.1.e describen la relación entre la segunda componente principal y los ángulos de azimut para los oídos izquierdo y derecho respectivamente. Sin embargo, y a pesar de que se pueden encontrar cierta relación entre los ángulos de azimut y la segunda componente principal, los coeficientes de correlación para los oídos izquierdo y derecho son bastante bajos (-0.0769 y 0.0016 respectivamente). Así también, si se comparan estas gráficas con las de la primera componente principal, se observa que presentan barras de error mayores. Por otra parte, la Figura 3.1.f, correspondiente a los ángulos de elevación, no presenta un patrón claro que sea capaz de describir los ángulos de elevación .

Con respecto a la tercera componente principal nos centraremos en la Figura 3.1.i, que evidencia una fuerte correlación entre los ángulos de elevación cercanos al plano medio y la tercera compone te principal. De hecho, de las seis componentes principales que se consideraron en el análisis, ésta es la que mayor coeficiente de correlación tiene con los ángulos de elevación (ver Tabla 3.1).

Hasta ahora hemos analizado la relación entre las tres primeras componentes principales y los ángulos de azimut y elevación. De las restantes componentes estudiadas no se puede concluir de manera directa sobre cómo influyen en la percepción del sonido del oyente, sin embargo, es importante destacar que estas componentes principales ayudan a mejorar la percepción de la ubicación de la fuente de sonido, ya que como vimos en la Sección 2.3

estas seis componente principales representan mas del 95 % de información original de las HRTFs.

Resumiendo, en la Figura 3.1 encontramos que el primer componente principal está fuertemente correlacionado con los ángulos de azimut y el tercer componte principal obtenido mediante PCA tiene cierta correlación con el ángulo de elevación. De forma general, para todos los componentes principales, la variabilidad entre sujetos aumenta en las ubicaciones ipsilateral y cuando las fuentes se alejan del plano frontal.

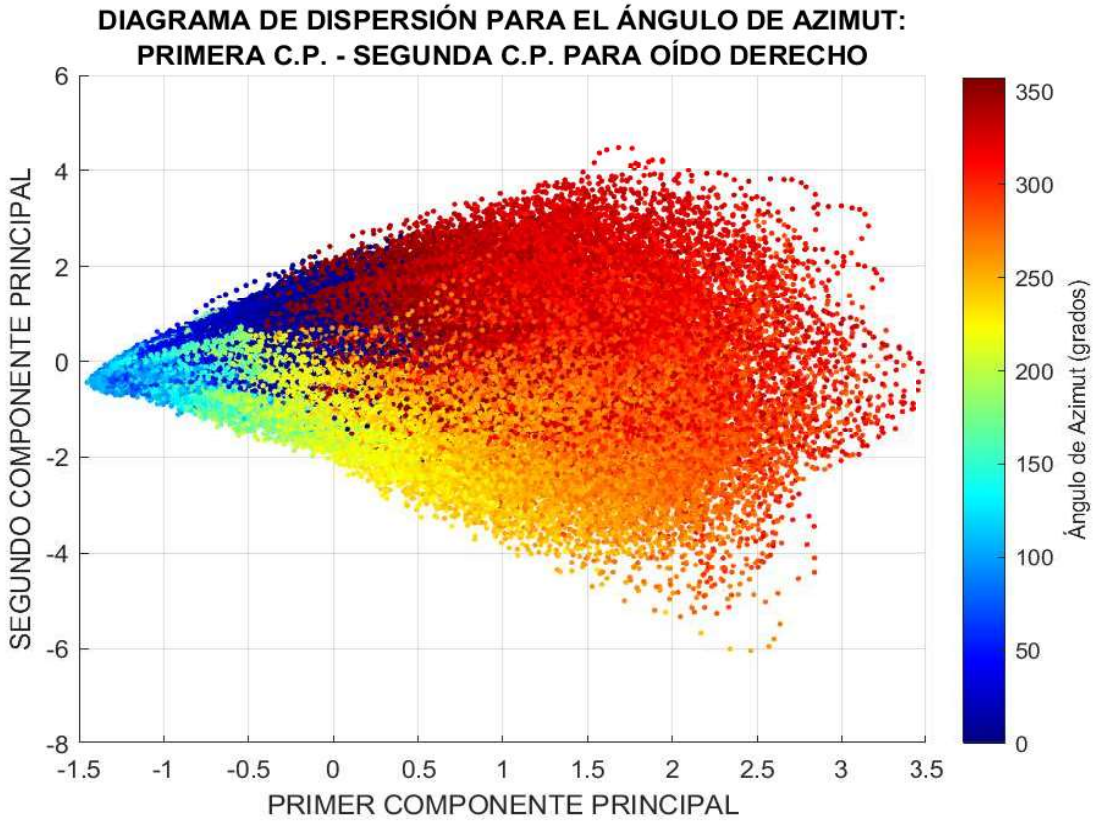

Figura 3.2: Diagrama de dispersión de los dos primeros componentes principales en función del azimut para el oído derecho.

La Figura 3.2 muestra cómo varían las dos primeras componentes principales con respecto al ángulo de azimut para el oído derecho, i.e. muestra cómo varía la primera componente frente a la segunda con respecto al ángulo de azimut, donde los ángulos de azimut han sido representados por una barra de color. En la figura se observa que existe una predominancia de colores rojos, esto debido a que la primera componente principal describe de mejor manera las fuentes de sonido ubicadas en el ipsilateral del oyente. Esto muestra que PCA captura de mejor manera la variabilidad del oído ipsilateral (i.e. el mas cercano a la fuente de sonido).

Por otra parte, se observa cómo los ángulos de azimut que corresponden a contralateral (i.e. ángulos menores a 180°) tienen poca presencia en la figura, lo cual indica que PCA no captura de forma adecuada la variabilidad del oído contralateral del oyente.

La Figura 3.3 muestra cómo varía la primera componente principal frente a la segunda con respecto al ángulo de azimut para el oído izquierdo. En esta figura se puede observar cómo a diferencia de la Figura 3.2, en esta figura existe un predominio de colores azules, los mismos que corresponden al ipsilateral del oyente (i.e. ángulos menores a 180°). Se observa también como los ángulos correspondientes al contralateral no son descritos de forma adecuada por su baja variabilidad.

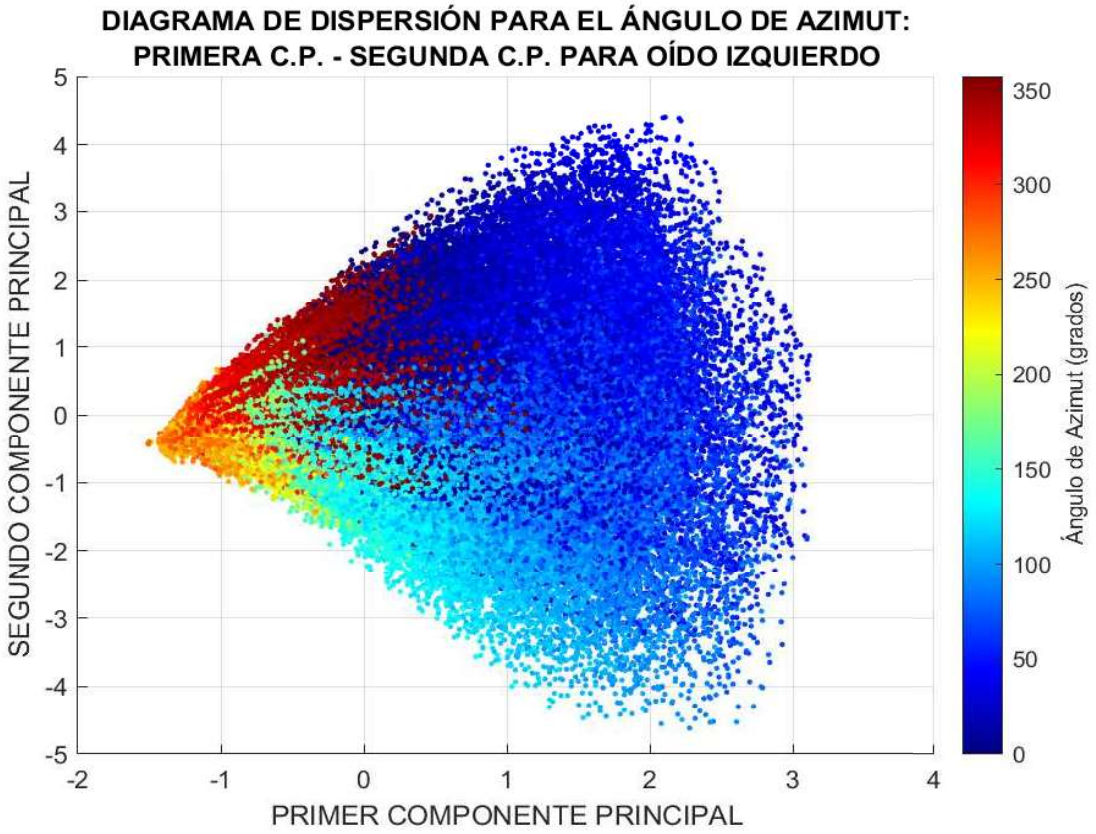

Figura 3.3: Diagrama de dispersión de los dos primeros componentes principales en función del azimut para el oído izquierdo

Ahora bien, si observamos la Tabla 3.1 comprobamos que no existe un valor de correlación significativo entre las primeras dos componentes principales y los ángulos de elevación. Esto es apreciable en la Figura 3.4 que muestra la relación entre la primera y la segunda componente principal para los ángulos de elevación, los mismos que están representados por una barra de color. En la figura no se puede apreciar un patrón claro que relacione las dos primeras componentes principales con el ángulo de elevación.

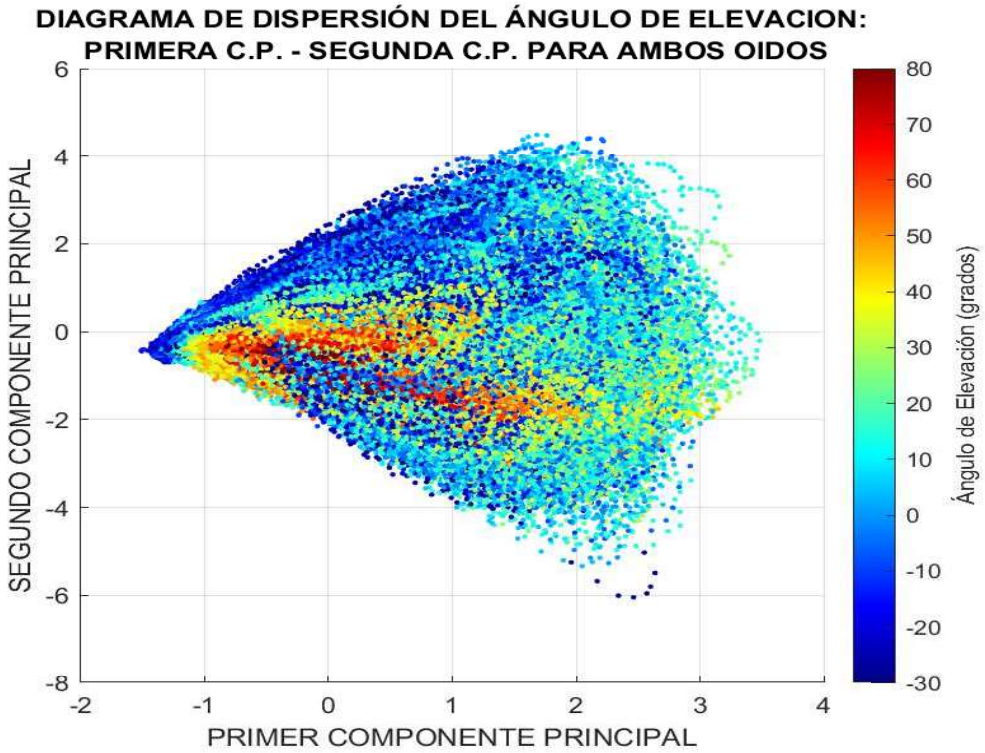

Figura 3.4: Diagrama de dispersión de los dos primeros componentes principales en función del ángulo de elevación para ambos oídos.

Resumiendo, las Figuras 3.2 y 3.3 muestran que las dos primeras componentes principales de las HRTFs obtenidas con PCA logran capturar la variabilidad del oído ipsilateral del oyente. Por otra parte, la Figura 3.4 muestra como estos mismos componentes principales no logran capturar mayor información de los ángulos de elevación.

3.2. RESULTADOS OBTENIDOS EN LA ETAPA DE SÍNTESIS

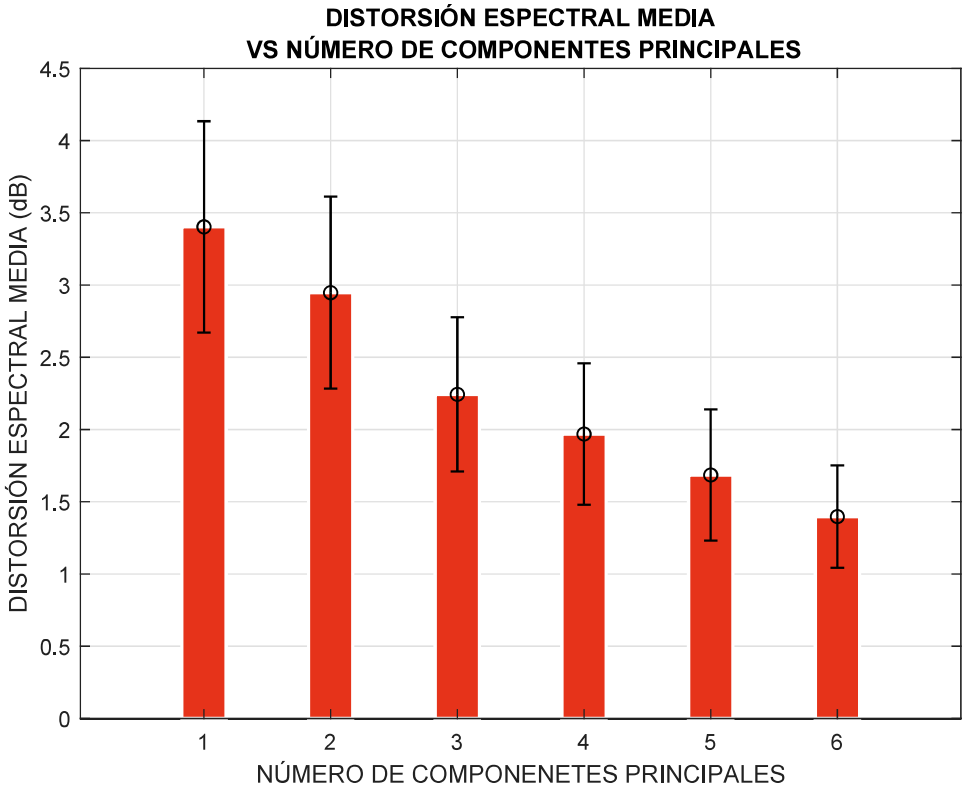

Figura 3.5: Distorsión espectral media en función del número de componentes principales retenidos. Las barras de error representan $\pm$ una desviación estándar.

La Figura 3.5 muestra el distorsión espectral media obtenido al reconstruir las funciones HRTFs de los 95 sujetos estudiados, usando de una hasta seis componentes principales. Tal y como se esperaba se observa como el valor de la distorsión espectral media disminuye a medida que aumentamos el número de componentes principales usados.

Además se observa que, en la Figura 3.5 al realizar la reconstrucción con una componente principal se obtuvo un error de $3{,}4024$ dB el cual disminuye a $1{,}397$ dB al usar las seis componentes principales. Fijándonos en la Tabla 3.2 vemos como al pasar de dos competentes principales a tres componentes se da la mayor reducción del valor de RMSE (esto si se comparan las componentes principales en parejas). Esto se puede explicar ya que las tres primeras componentes principales obtenidas con PCA son las que tienen los coeficientes de correlación mas altos según la Tabla 3.1.

La Figura 3.5 muestra también como la desviación estándar disminuye a medida que aumentan el número de componentes principales.

Continuando con la Figura 3.6 que muestra la distorsión espectral media de cada sujeto para todas las posiciones de la fuente de sonido (i.e. considerando todos los ángulos de azimut y elevación) con seis componente principales, tenemos que la distorsión espectral media de 1.3971 dB con una desviación estándar de 0.3543.

Tabla 3.2: Distorsión espectral media

Número de Componentes Principales	Distorsión espectral media	Desviación estándar
1	3.4024	0.7317
2	2.9475	0.6645
3	2.2432	0.5342
4	1.9688	0.4895
5	1.6851	0.4537
6	1.3971	0.3543

Por otra parte, la Figura 3.7 muestra la distorsión espectral resultado de la síntesis de funciones HRTFs usando diferente número de componentes principales.La Figura 3.7.a muestra como varia la distorsión espectral con respecto a la frecuencia para los 95 sujetos estudiados, reconstruyendo las funciones HRTFs con una componente principal. Se observa como en este caso existe una mejor reconstrucción de bajas frecuencia. Mientras que las de frecuencias altas (i.e. superiores a 6 kHz) presentan una distinción espectral significativamente mayor al de las bajas frecuencias. Esta tendencia se mantiene al usar dos, tres, cuatro y cinco componentes principales.

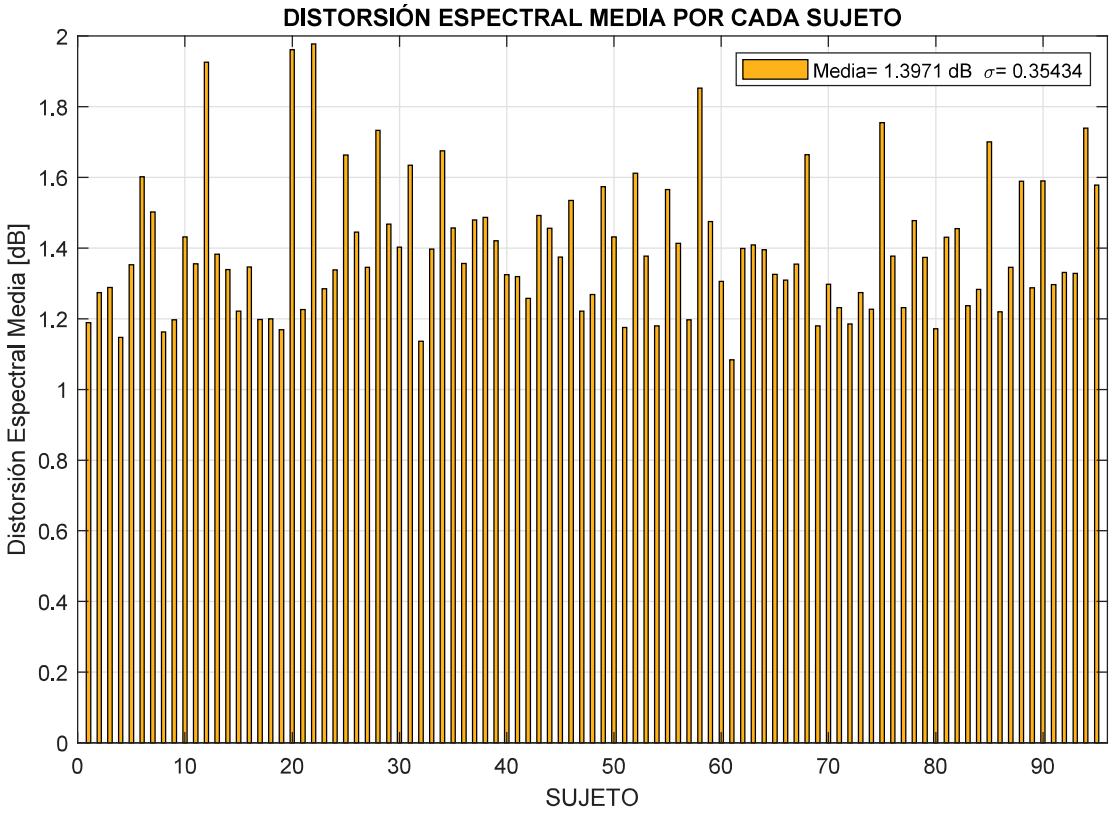

Figura 3.6: Distorsión espectral media por sujeto.

Se observa también que, las Figuras 3.7.c, 3.7.d y 3.7.f mantienen un patrón presente en la Figura 3.7.a, pero además es apreciable como en estas figuras se presentan un nuevo

patrón, en el que las componentes de frecuencia comprendidas entre 4000 Hz y 7000 Hz presentan una menor distorsión espectral. Por otra parte, en la Figura 3.7.f, que corresponde a la distorsión espectral de las funciones HRTFs usando seis componentes principales, se aprecia cómo el patrón descrito anteriormente cambia, y ahora tenemos que frecuencias mayores a 5000 Hz pero menores a 8000 Hz presentan una mejor reconstrucción de las funciones HRTFs (ya que tienen una menor distorsión espectral). Mientras que frecuencias menores a 5000 Hz y mayores a 8000 Hz tienen una distorsión espectral mayor.

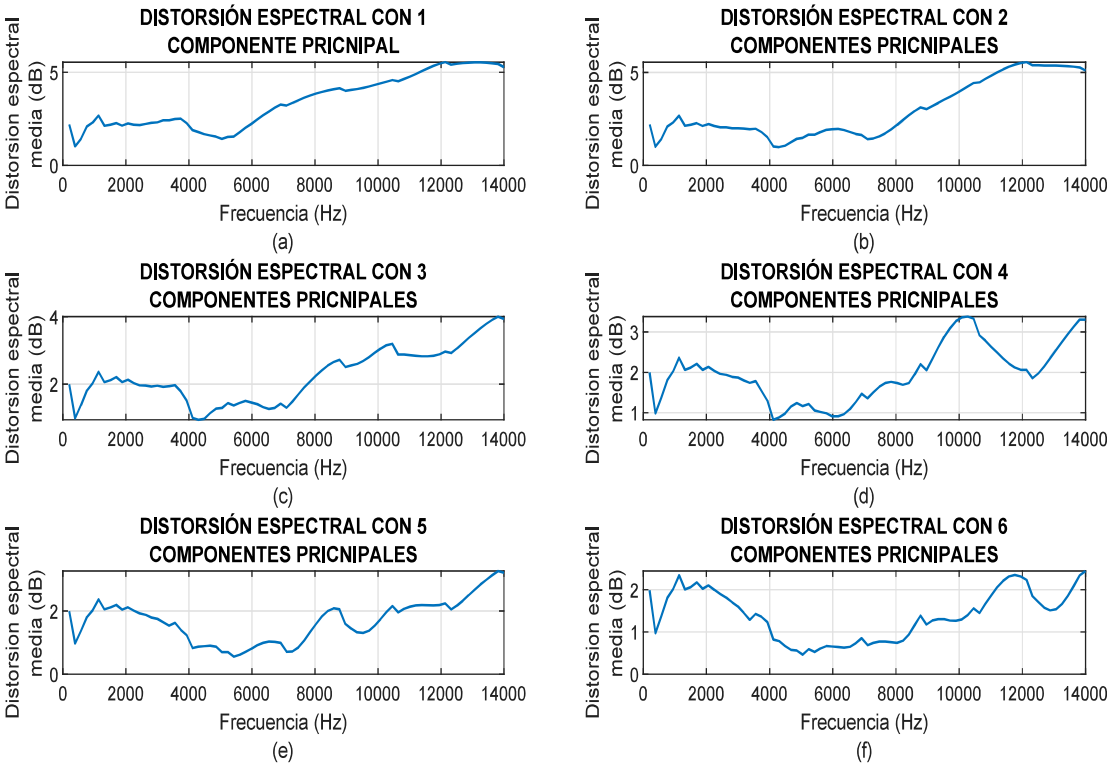

Figura 3.7: Variación de distorsión espectral a lo largo de las seis primeras componentes principales

En la Figura 3.8, se muestra la distorsión espectral media con respecto a los ángulos de azimut para los oídos izquierdo y derecho, obtenida al realizar la síntesis de funciones HRTFs usando PCA con seis componentes principales. Para el cálculo de la distorsión espectral media se han considerado todos los ángulos de elevación presentes en la base de datos. La Figura 3.8.a presenta la distorsión espectral para el oído izquierdo, donde se observa que existe una baja distorsión en ángulos de azimut correspondientes al ipsilateral del oyente. Por otra parte, los ángulos correspondientes al contralateral van incrementando su distorsión espectral hasta alcanzar una distorsión máxima de 2.914 dB en 250°. Entre 250° y 280° la distorsión espectral baja nuevamente para encontrar otro pico en 295°. La Figura 3.8.b muestra una relación similar pero para el oído derecho, en la que encontramos la máxima distorsión espectral en 110° con un valor de 2.793 dB. La presencia de dos “picos” en las Figuras 3.8.a y 3.8.b puede ser explicada por la presencia de reflexiones y refracciones en la cabeza del oyente.

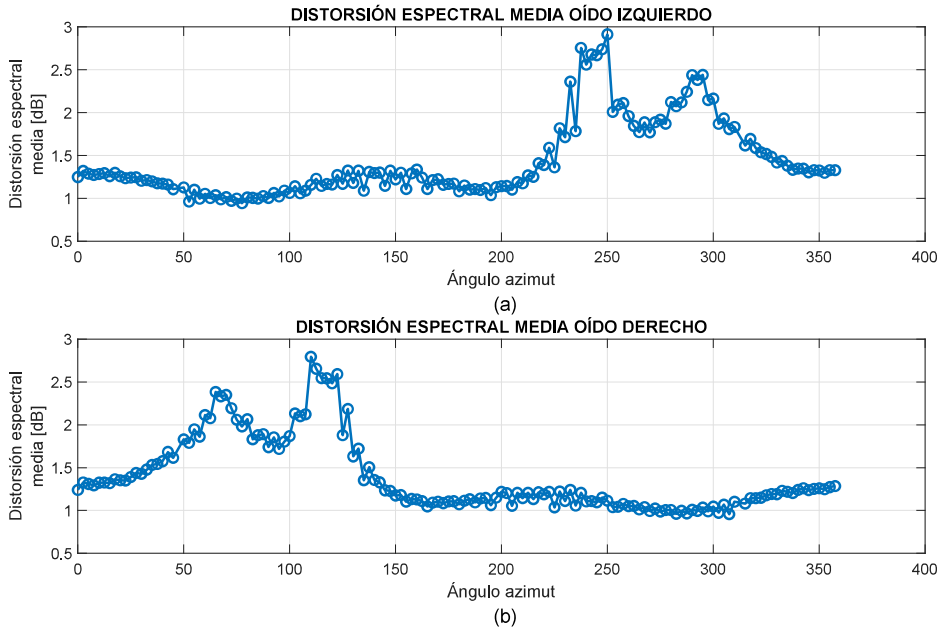

Figura 3.8: Distorsión espectral media para para los oídos izquierdo y derecho con seis componentes principales.

Por otra parte, en la Figura 3.9, se muestra la distorsión espectral media con respecto a los ángulos de elevación para los oídos izquierdo y derecho, obtenida al realizar la síntesis de funciones HRTFs usando PCA con seis componentes principales. Para el cálculo de la distorsión espectral media se han considerado todos los ángulos de azimut presentes en la base de datos. En al figura se observa como, la distorsión espectral disminuye a medida que la fuente de sonido se mueve hacia el plano frontal (i.e. $\phi = 90°$) alcanzando un valor mínimo de 0.9556 dB en 80 ° para el oído derecho y 0.9518 dB en 80 ° para el oido izquierdo.

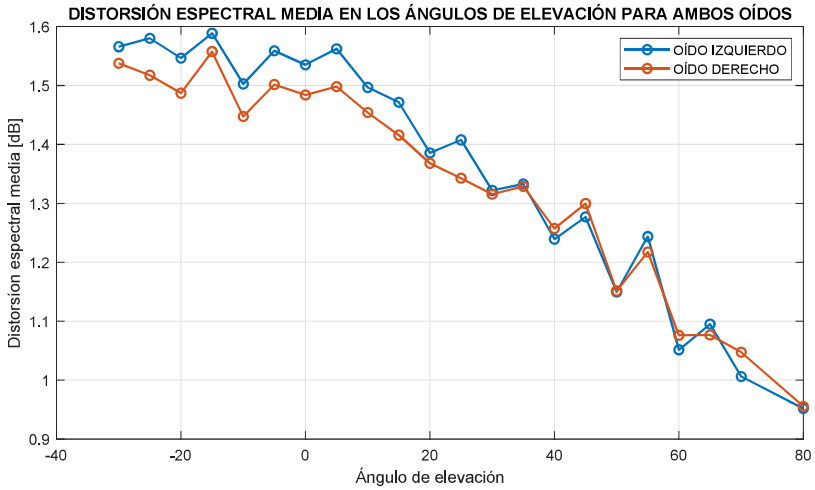

Figura 3.9: Distorsión espectral media para para los oídos izquierdo y derecho con seis componentes principales.

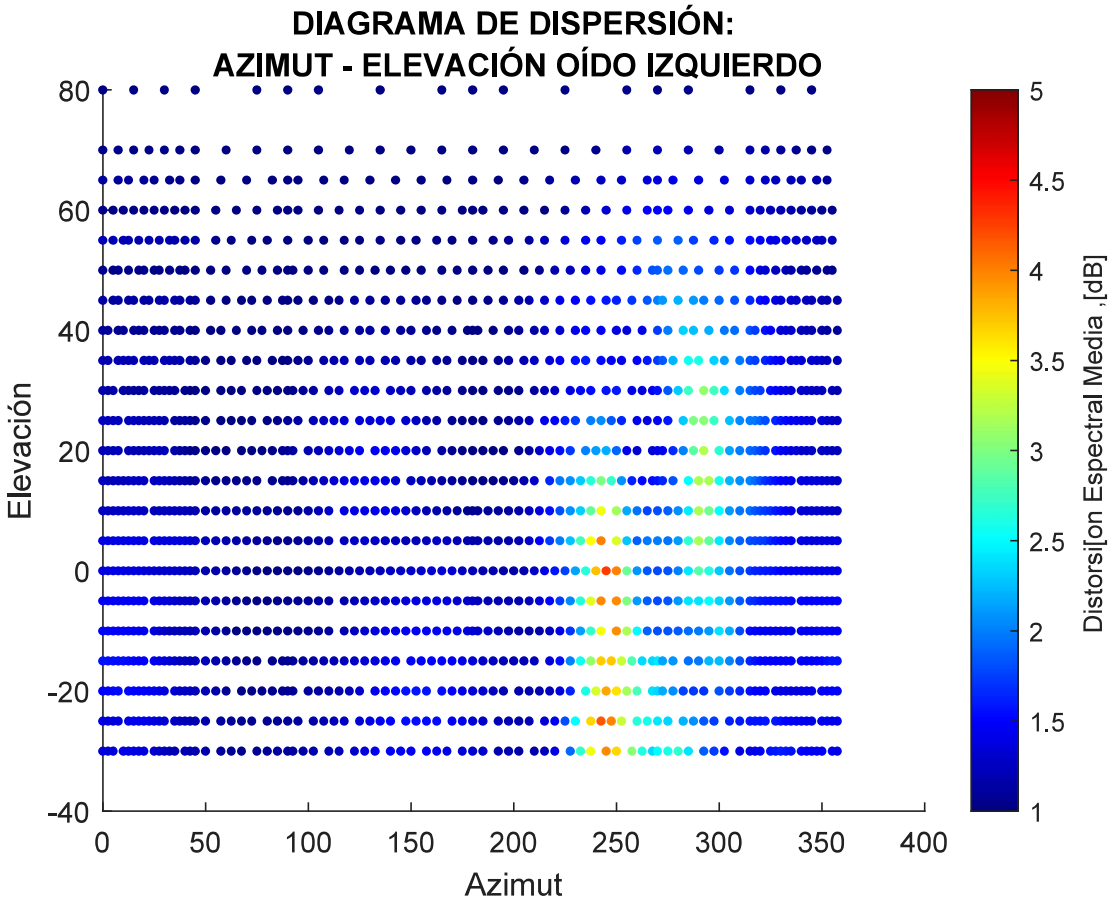

Figura 3.10: Diagrama de dispersión para el oído izquierdo.

En la Figura 3.10 se muestra la distorsión espectral media obtenida al sintetizar las funciones HRTFs con seis componentes principales con respecto a los ángulos de elevación y azimut. Para esto se uso una barra de color. En la figura observamos que existe mayor distorsión espectral en ángulos de azimut cercanos a los 250° y también cercanos a plano medio del oyente (i.e cerca a $\theta = 0°$). Se aprecia también que existe una distorsión espectral significativa para ángulos de azimut cercanos a 300° y que la máxima distorsión espectral esta en los ángulos 0 ° y 245 ° de elevación y azimut respectivamente con un valor de 4.2627 dB.

Se observa también que en ángulos correspondientes al ipsilateral del oyente (i.e. ángulos cercanos a $\theta = 90°$) la distorsión espectral media es significativamente menor con valores cercanos a 1.5 dB.

La Figura 3.11 muestra la distorsión espectral media obtenida al sintetizar las funciones HRTFs con seis componentes principales con respecto a los ángulos de elevación y azimut. Para esto se uso una barra de color. En la figura observamos que existe mayor distorsión espectral en ángulos de azimut cercanos a los 110° y también cercanos a plano medio del oyente (i.e cerca a $\theta = 0°$). Se aprecia también que existe una distorsión espectral significativa para ángulos de azimut cercanos a 70° y que la máxima distorsión espectral esta en los ángulos 0 ° y 155 ° de elevación y azimut respectivamente con un valor de

3.8695 dB.

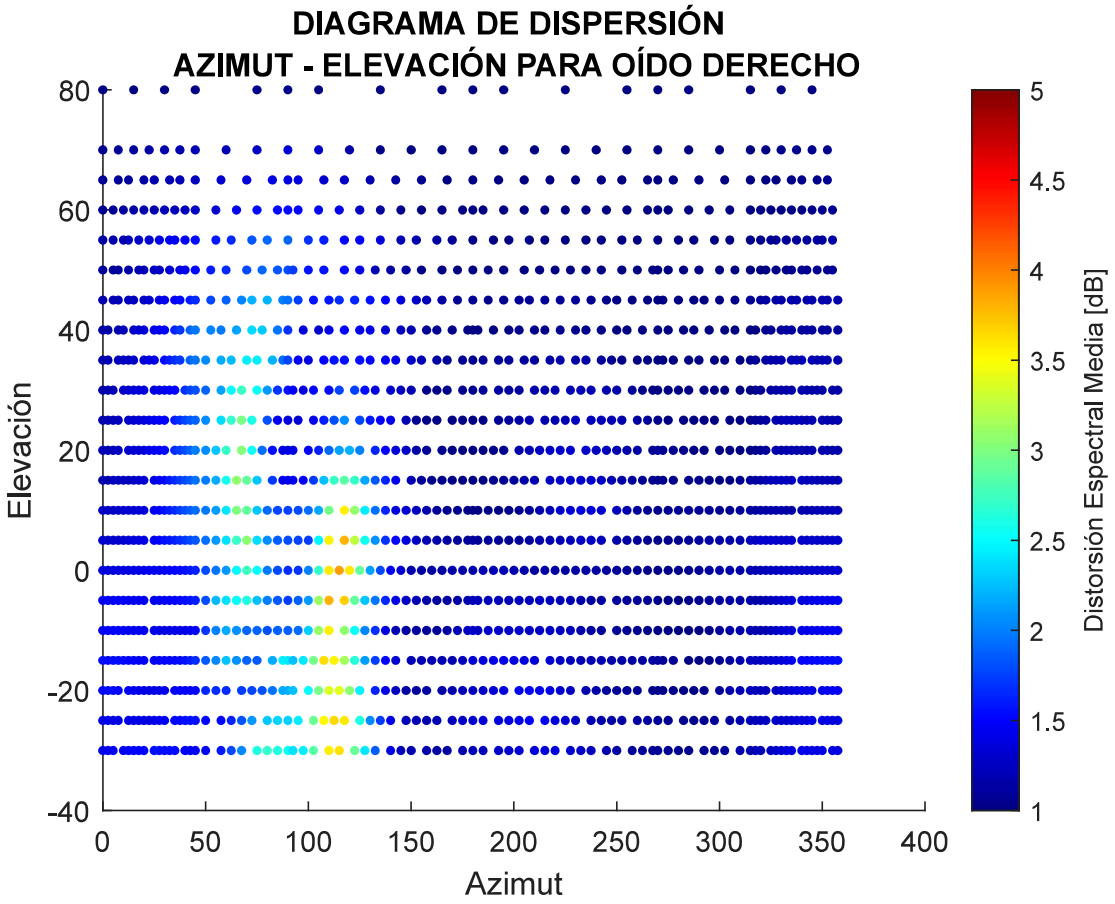

Figura 3.11: Diagrama de dispersión para el oído izquierdo.

De igual manera observamos en la figura que en ángulos correspondientes al ipsilateral del oyente (i.e. ángulos cercanos a $\theta = 180°$) la distorsión espectral media es significativamente menor con valores cercanos a 1.5 dB.

4. CONCLUSIONES Y RECOMENDACIONES

4.1. CONCLUSIONES

En el presente proyecto se abordaron tres etapas, pre-procesamiento, análisis y síntesis, sin embargo, las conclusiones se centrarán en las dos últimas etapas del mismo ya que estas presentan el aporte más significativo.

Dentro del investigación bibliográfica realizada, nos centramos esencialmente en el estudio de audio espacial y la aplicación de PCA. Se ha podido concluir que el análisis de componentes principales al ser una técnica utilizada para presentar un conjunto de datos en términos de variables nuevas (i.e componentes principales) puede ser utilizada para reducir dimensionalidad en un conjunto de datos. Esto hace de PCA una herramienta potente al momento de analizar características espectrales de las funciones HRTFs, las mismas que son fundamentales en el estudió de audio espacial. Del estudio de las HRTFs se concluyen que estas funciones describen los patrones de directividad de los oídos.

Dentro de la etapa de análisis, se redujo la dimensionalidad de la base de datos original (i.e. aquella obtenida después de haber sido filtrada y suavizada espectralmente) de 75 componentes (o variables) originales a seis componentes principales luego de aplicar PCA, reduciendo así considerablemente la dimensionalidad de los datos.

Ahora bien, una vez seleccionado el número de componentes principales para realizar la síntesis, se encontró que las primeras tres competentes principales obtenidas con PCA son las que describen de mejor manera los ángulos de azimut y elevación. Específicamente se encontró que la primera componente principal esta fuertemente correlacionada y por lo tanto describe de mejor manera los ángulos de azimut. En contraste, la tercera componente principal describe de forma pobre los ángulos de elevación, sobre todo si se la compara con la primera componente principal. Sin embargo, es la componente principal que presenta mayor correlación con los ángulos de elevación.

Por otra parte las restantes componentes principales no describen de manera directa los ángulos de azimut y elevación, sin embargo, se puede concluir que influyen en la percepción del sonido del oyente ya que en conjunto las seis componentes principales representan mas del 95 % de la información de las HRTFs originales. Además, si consideramos los diagramas de dispersión de la etapa de análisis se observa que en la primera y segunda componente principal los ángulos de azimut tiene mayor variabilidad en el ipsilateral del oyente, esto hace que la información que llega a los oídos desde el contralateral del oyente sea pobremente reconocida por las primeras componentes principales.

Finalmente dentro de la etapa de síntesis, se puede concluir que las figuras que involucran a los ángulos de azimut describen simetría en los oídos izquierdo y derecho, es decir describen simetría con respecto al plano medio del oyente. Con respecto a la distorsión espectral media concluimos que la misma aumenta a medida que nos acercamos al contralateral del

oyente, lo cual es esperado debido a que la cabeza del sujeto actúa ejerciendo un efecto de sombra [1] sobre el sonido haciendo que la síntesis sea más compleja. Por ultimo, cabe mencionar que la distorsión espectral media es relativamente baja, considerando que ha logrado capturar el 95 % de varianza, lo que implica que PCA captura información relevante de las HRTFs de manera adecuada.

4.2. RECOMENDACIONES

Una vez concluido este proyecto consideramos que seria interesante ampliar el estudio en el campo del audio espacial, por lo tanto se recomienda aplicar el análisis de componentes principales a funciones HRTFs de campo cercano, y comparar los resultados obtenidos con los expuestos en este proyecto.

Así mismo, dentro de este campo de estudio se recomienda aplicar SPCA (*Spatial Principal Component Analysis*) [1] a funciones HRTFs de campo lejano y cercano.

También se recomienda para trabajos futuros, analizar técnicas de reducción de dimensionalidad no lineal, e.g., Isomap [37].

Finalmente, como recomendación resultante de este trabajo se recomienda, tomar en cuenta que al hacer *k-fold cross-validation* sobre funciones HRTFs, la aleatorización se la debe realizar a lo largo de los sujetos para evitar contaminación de datos.

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] H. MøllеR, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, pp. 171–218, 12 1992.

[3] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[4] 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," May 2013.

[5] V. Autores, "Oculus: 3d audio spatialization," citado el 22/01/2019. [Online]. Disponible en: https://developer.oculus.com/documentation/audiosdk/latest/concepts/book-audio-intro/

[6] A. Kohlrausch, J. Braasch, D. Kolossa, and J. Blauert, *An Introduction to Binaural Processing*, 2013, pp. 1–32.

[7] J. C Schacher and P. Kocher, "Ambisonics spatialization tools for max/msp," 01 2006.

[8] C. D. Geisler, *From Sound to Synapse: Physiology of the Mammalian Ear*. Oxford University Press, 1998. [Online]. Disponible en: https://books.google.com.ec/books?id=baRutmLuDJsC

[9] F. P. Freeland, L. Biscainho, and P. Diniz, "Interpositional transfer function for 3d-sound generation," *Journal of the Audio Engineering Society*, vol. 52, pp. 915–930, 09 2004.

[10] B. Moore, *An Introduction to the Psychology of Hearing*. Emerald, 2012. [Online]. Disponible en: https://books.google.com.ec/books?id=LM9U8e28pLMC

[11] L. F. Wightman and D. 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, 1999.

[12] P. Stephen and N. William, "The efect of head rotations on vertical plane sound localization," *The Journal of the Acoustical Society of America*, vol. 105, no. 4, pp. 2325–2332, 1997.

[13] J. Holzl, "A global model for hrtf individualization by adjustment of principal component weights," Ph.D. dissertation, kunstunigraz, 2014.

[14] H. Gamper, “Head-related transfer function interpolation in azimuth, elevation, and distance,” *The Journal of the Acoustical Society of America*, vol. 134, no. 6, pp. EL547–EL553, 2013.

[15] M. C. G. Ramos, “Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications,” *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 1735–1738, 2013.

[16] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki, and T. Sone, “A database of head-related transfer functions in whole directions on upper hemisphere,” *Acoustical Science and Technology*, vol. 23, pp. 160–162, 05 2002.

[17] 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, pp. 623–637, 07 2007.

[18] D. G. Stan, J.J. Embrecht, “Comparison of different impulse response measurement techniques,” *Journal of the Audio Engineering Society*, vol. 50, pp. 249–262, 04 2002.

[19] C. Oreinos and J. Buchholz, “Measurement of a full 3d set of hrtfs for in-ear and hearing aid microphones on a head and torso simulator (hats),” *Acta Acustica united with Acustica*, vol. 99, 09 2013.

[20] “Catalog of American National Standards on Acoustics,” *The Journal of the Acoustical Society of America*, vol. 87, no. 1, pp. 457–460, Jan. 1990. [Online]. Disponible en: https://asa.scitation.org/doi/abs/10.1121/1.399240

[21] B. K. Head, “45bc KEMAR Head and Torso with Mouth Simulator,” p. 17.

[22] 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.

[23] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.

[24] E. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” *The Journal of the Acoustical Society of America*, vol. 94, pp. 111–23, 08 1993.

[25] H. Geoffrey Fisher and S. J. Freedman, “The role of the pinna in auditory localization,” *Journal of Auditory Research*, vol. 8, pp. 15–26, 01 1968.

[26] P. Balazs, R. Abbasi, and L. D. Abreu, “Ari hrtf database,” https://www.kfs.oeaw.ac.at, 2019.

[27] V. A. Oppenheim and W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.

[28] 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, 1992.

[29] S. K. I. A. Kulkarni and H. S. Colburn., “On the minimum-phase approximation of head-related transfer functions,” *In IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 85–94, 1995.

[30] A. Y.-T. Ng, “Machine learning,” https://www.coursera.org/learn/machine-learning, Feb. 2019. [Online]. Disponible en: https://www.coursera.org/

[31] G. Ivosev, L. Burton, and R. Bonner, “Dimensionality Reduction and Visualization in Principal Component Analysis,” *Analytical Chemistry*, vol. 80, no. 13, pp. 4933–4944, Jul. 2008. [Online]. Disponible en: https://pubs.acs.org/doi/10.1021/ac800110w

[32] R. J. Goult, *Applied linear algebra*, ser. Mathematics & its applications. Ellis Horwood, 1978. [Online]. Disponible en: https://books.google.com.ec/books?id=eAHvAAAAMAAJ

[33] T. Tr, “Dimensionality Reduction: A Comparative Review,” p. 36.

[34] F. Grijalva, “Manifold learning for spatial audio rendering (aprendizado de variedades para a sintese de audio espacial),” Ph.D. dissertation, Cottbus, 2018.

[35] P. Balazs, R. Abbasi, and L. D. Abreu, “Sofa (spatially oriented format for acoustics),” http://www.sofaconventions.org/, 14 May 2013.

[36] D. F. y. S. G. F. Grijalva, L. C. Martini, “Interpolation of head-related transfer functions using manifold learning,” *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 221–225, Feb 2017.

[37] M. Balasubramanian and E. L. Schwartz, “The isomap algorithm and topological stability,” *Science*, vol. 295, no. 5552, pp. 7–7, 2002. [Online]. Disponible en: https://science.sciencemag.org/content/295/5552/7

ANEXOS

ANEXO A. main_f.m

ANEXO B.Función validar_datos

ANEXO C. Función extraccion_datos.m

ANEXO D. Función indices.m

ANEXO E. Función vector_aleatorio.m

ANEXO F. K-FOLD CROSS-VALIDATION

ANEXO G. Función filtrado_HRTF.m
(El código se encuentra en el CD adjunto a este documento)

ANEXO H.Función filtrado_HRTF2.m
(El código se encuentra en el CD adjunto a este documento)

ANEXO I. Función GRAFICAS_K_FOLD.m
(El código se encuentra en el CD adjunto a este documento)

ANEXO J.ObtencionDatosGraficos.m
(El código se encuentra en el CD adjunto a este documento)

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

ANEXO A: main_f.m

```
%%INICIO
SOFAstart;
HRTF=SOFAload('C:\Users\USUARIO\BORIS_POLI\TESIS-PCA\API_MO-master\HRTFs...
\ARI_HRTF\hrtf_nh164.sofa');
Fs=HRTF.Data.SamplingRate;
N=size(HRTF.Data.IR,3);
Muestras=2*size(HRTF.Data.IR,1);
%%EXTRAER DATOS
path='C:\Users\USUARIO\BORIS_POLI\TESIS-PCA\API_MO-master\HRTFs\ARI_HRTF';
extension='.sofa';
datos_validados=validar_datos(path,extension);
TABLA_NUEVA=extraccion_datos(path,datos_validados,Muestras);
save('DATOS_ORIGINALES.mat','TABLA_NUEVA')
%%APLICACION DE LA TRANSFORMADA RAPIDA DE FOURIER
HRIR=TABLA_NUEVA.HRIR;
 %Se aplica la funcion fft a lo largo de cada fila
HRTF_x_two_slided=abs(fft(HRIR,[],2));
%Se toma unicamente la mitad de los datos de la fft
HRTF_x_single_slided=HRTF_x_two_slided(:,1:floor(N/2)+1);
 %Se multiplica todos los valores por dos excepto la componente DC
HRTF_x_single_slided(:,2:end-1)=2*HRTF_x_single_slided(:,2:end-1);
%Se agrega a la tabla la colunma HRTF
TABLA_NUEVA.HRTF=HRTF_x_single_slided;
TABLA_HRTF=TABLA_NUEVA;
save('DATOS_HRTF.mat','TABLA_HRTF')
%%FILTRADO Y SUAVIZADO ENTRE 200 Hz y 14000Hz
DATOS=struct('fmin',200,'fmax',14000,'Fs',48000);
 %Funcion de suavizado espectral
HRTF_SUAVIZADO=HRTFsmoothing(TABLA_HRTF.HRTF,48000);
%Funcion de filtrado espectral
HRTF_FILTRADO=filtrado_HRTF(HRTF_SUAVIZADO,DATOS);
%Se agrega la columna de HRTF_FILTRADO_SUAVIZADO
TABLA_HRTF.HRTF_FILTRADO_SUAVIZADO=HRTF_FILTRADO;
 %Armado de nueva tabla de datos
BASE_DATOS=TABLA_HRTF;
BASE_DATOS.HRIR=[];
BASE_DATOS.HRTF=[];
%Se guarda la nueva tabla
save('DATOS_PROCESADOS.mat','BASE_DATOS')
%%APLICANDO PCA SOBRE LA BASE DE DATOS
[coeff,score,varianzas,¬,¬,mu]=pca(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO);
save('DATOS_ANALIS.mat')
%%FUNCIONES HRTF EN dB
HRTFdB=20*log10(abs(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO));  %FUNCION HRTF EN dB
%%RECONSTRUCCION
n_componentes=6;
for j=0:n_componentes-1
```

```
HRTFapprox(j*size(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO,1)+1:(j+1)*...
      size(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO,1),:)...
      =score(:,1:j+1)*coeff(:,1:j+1)';
end
% Suma de la mediaa las funciones HRTFs
HRTFapprox=HRTFapprox+repmat(mu,size(HRTFapprox,1),1);
%%FUNCIONES HRTFs APROXIMADAS EN dB
HRTFapproxdB=20*log10(abs(HRTFapprox));
%%CALCULO DE DISTORSION ESPECTRAL Y ERROR CUADRATICO MEDIO
% Calculo de distorsión espectral
SD = abs(abs(repmat(HRTFdB,n_componentes,1))-abs(HRTFapproxdB));
RMSE = sqrt(mean(SD.^2,2)); %CALCULO DEL ERROR CUADRATICO MEDIO
VAR = var(SD,0,2);      %CALCULO DE LA VARIANZA POR FILA
%%NUEVA TABLA
%SE INCLUYE COLUMNAS DE DISTORSION ESPECTRAL, RSME Y VAR
for j=0:n_componentes-1
    NUMERO_COMPONENTES(j*size(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO,1)+1:(j+1)...
        *size(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO,1),:)...
        =(j+1)*ones(size(BASE_DATOS.HRTF_FILTRADO_SUAVIZADO,1),1);
end
nueva=BASE_DATOS{:,:};
nueva=repmat(nueva,n_componentes,1);
NUEVA_TABLA=table(nueva(:,1),nueva(:,2),nueva(:,3),nueva(:,4),nueva(:,5:79),...
    'VariableNames',{'SUJETO','AZIMUT','ELEVACION','EAR',...
    'HRTF_FILTRADA_SUAVIZADA'});
TABLA=table(NUMERO_COMPONENTES,HRTFapprox,SD,RMSE,VAR);
DATOS=[NUEVA_TABLA,TABLA];
```

ANEXO B: Función validar_datos

```
function lista = validar_datos(path,extension)
%path                ruta de la carpeta, si es la actual  poner path=pwd
%extension           extension(formato), si no se desea filtrar por extension ...
    poner ext=''
lista_archivos=ls(path);
k=1;                  %contadores auxilires
 for j=1:size(lista_archivos,1)
   aux=lista_archivos(j,:);      %Se lee la linea 'j' de lista_archivo
   [¬,¬,ex]=fileparts(aux);      %Se toma la extension de los archivos
    %Scomprueba la direccion de path es valida y se extrae solo os archivos ...
        '.sofa'
   if (and(¬isfolder(fullfile(path,aux)),or(strcmpi(strtrim(ex),extension),...
   isempty(extension))))
       hrtf=SOFAload(fullfile(path,aux));    %Se carga el archivo '.sofa'
       muestras=length(hrtf.Data.IR);        %Numero de muestras
      if muestras == 1550
         lista_bien(k)=j;
         k=k+1;
      end

   end
 end
 lista=lista_archivos(lista_bien,:);
end
```

ANEXO c: Función extraccion_datos.m

```
function tabla = extraccion_datos(path,lista_archivos,numero_muestras)
tamanio=numero_muestras;      %NUMERO DE MUESTRAS POR INDIVIDUO
for j=0:size(lista_archivos,1)-1
    aux=lista_archivos(j+1,:);                %SE LEE LA LINEA 'j' de ...
        lista_archivo
    hrtf=SOFAload(fullfile(path,aux));        %SE CARGA EL ARCHIVO *.sofa
    %SE EXTREA LA MATRIZ DE IR
    HRIR(j*tamanio+1:(j+1)*tamanio,:)=...
    [squeeze(hrtf.Data.IR(:,1,:));squeeze(hrtf.Data.IR(:,2,:))];
    %SE EXTREA LAMMATRIZ DE AZIMUT
    AZIMUT(j*tamanio+1:(j+1)*tamanio,:)=...
    [hrtf.SourcePosition(:,1);hrtf.SourcePosition(:,1)];
    %SE EXTREA LAMMATRIZ DE ELEVACION
    ELEVACION(j*tamanio+1:(j+1)*tamanio,:)=...
    [hrtf.SourcePosition(:,2);hrtf.SourcePosition(:,2)];
    %SE CREA LA MATRIZ DE OIDOS
    EAR(j*tamanio+1:(j+1)*tamanio,:)=[ones(tamanio/2,1);2*ones(tamanio/2,1)];
    %SE AGRAGA EL ID DEL SUJETO
    SUJETO(j*tamanio+1:(j+1)*tamanio,:)=(j+1)*ones(tamanio,1);
end
 tabla=table(SUJETO,AZIMUT,ELEVACION,EAR,HRIR);
end
```

ANEXO D: Función indices.m

```
function [sujetoEindice,indice]=indices(Matriz,NumeroClases,TamanioClase )
%Matriz          La matriz de datos debe ingresar como una matriz con las ...
    observacines en las filas
%NumeroClases   Indica en cuantos grupos se dividira la matriz
%TamanioClase   El numero de muestras correspondiente a cada usuario (parta ...
    las HRTF 3100)
NumeroSujetos=size(Matriz,1)/TamanioClase;
tamanio=NumeroSujetos/NumeroClases;
aux1=1:NumeroClases;
aux2=vector_aleatorio(NumeroSujetos);
aux3=repmat(aux1,tamanio,1);
aux4=repmat(aux2,TamanioClase,1);
aux5=reshape(aux4,size(Matriz,1),1);
indice=ceil(NumeroClases*(aux5/NumeroSujetos));
sujetoEindice=[aux5,indice];
end
```

ANEXO E: Función vector_aleatorio.m

```
function m = vector_aleatorio(tamanio)
n=0;                %CANTIDAD DE NUMEROS ALEATORIOS VALIDOS

while n<tamanio
    a=randi(tamanio);       %GENERA UN ENTERO AL AZAR
    i=0;                %INDICE DEL VECTOR
    bandera=0;          %BANDERA QUE DETERMINA SI DOS NUMEROS SON IGUALES (0 SI ...
        NO HAY UN NUMERO IGUAL, 1 EN CASO CONTRARIO)
    while i<n && bandera==0
        i=i+1;
        if m(i)==a
            bandera=1;          %LA BANDERA CAMBIA SI EN NUMERO ALEATORIO, YA A ...
                SIDO USADO
        end
    end
    if bandera==0
        n=n+1;              %SE INCREMENTA LOS NUMEROS ALEATORIOS VALIDOS
        m(n)=a;             %SE ALMACENA EL NUMERO a EN LA POSION n
    end
end
end
```

ANEXO F: K-FOLD CROSS-VALIDATION

```
%%SEGUNDO EXPERIMENTO
%%OBTENCION DE DATOS PARA EL EXPERIMENTO
DatosValidacion=DATOS(DATOS.NUMERO_COMPONENTES==6,{'SUJETO','AZIMUT'...
    ,'ELEVACION','EAR','HRTF_FILTRADA_SUAVIZADA'});
DatosValidacion.Properties.UserData='NUMERO DE COMPONENTES FIJADO EN 6 ...
    COMPONENTES PRINCIPALES';
save('DATOS_para_K_FOLD.mat','DatosValidacion')
%%PREPARACION DE DATOS K-FOLD CROSS VALIDATION
load DATOS_para_K_FOLD.mat
folds=5;          %Numero de Grupos en los que se divide los datos originales

%Función indices: devuelve las matrices sujetos  e indices, donde sujeto es ...
    una matriz  que cuenta con los sujetos y a que indice pertenecen,...
indice solo muestra los indices
[sujetos,indice]=indices(DatosValidacion.SUJETO,folds,3100);

 save('DATOS_para_K_FOLD.mat','DatosValidacion','sujetos','indice','fold')
%%K-FOLD CROSS VALIDATION
load DATOS_para_K_FOLD.mat
longitud_test=length(DatosValidacion.SUJETO)/fold; %numero de sujetos de ...
    cada fold
numero_componentes=6; % componentes pricnipales
for j = 1:numero_componentes
for i = 1:fold
    test = (indice == i);
    train = ¬test;
    mat_test((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.HRTF_FILTRADA_SUAVIZADA(test,:);
    mat_azimut((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.AZIMUT(test,:);
    mat_elevacion((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.ELEVACION(test,:);
    mat_ear((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        DatosValidacion.EAR(test,:);
    [MATRIZ_ROTACIONAL,¬,¬,¬,¬,MEDIA]=...
    pca(DatosValidacion.HRTF_FILTRADA_SUAVIZADA(train,:));
    PROYECCIONES=(DatosValidacion.HRTF_FILTRADA_SUAVIZADA(test,:)-...
        MEDIA)*MATRIZ_ROTACIONAL(:,1:j);
    RECONSTRUCCION=PROYECCIONES*MATRIZ_ROTACIONAL(:,1:j)'+ MEDIA;
    mat_recon((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=RECONSTRUCCION;
    folds((i-1)*longitud_test+1:((i-1)+1)*longitud_test,:)=...
        i*ones(longitud_test,1);
end

SD_k_fold = abs(20*log10(abs(mat_test))-20*log10(abs(mat_recon)));

SUJETOS((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)...
```

```
    *length(DatosValidacion.SUJETO),:)=sujetos(:,1);
FOLD((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)...
    *length(DatosValidacion.SUJETO),:)=folds;
COMPONENTES_PRINCIPALES((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)...
    *length(DatosValidacion.SUJETO),:)=j*ones(length(DatosValidacion.SUJETO),1);
AZIMUT((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)*...
    length(DatosValidacion.SUJETO),:) =mat_azimut;
ELEVACION((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)*...
    length(DatosValidacion.SUJETO),:)=mat_elevacion;
EAR((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)...
    *length(DatosValidacion.SUJETO),:)=mat_ear;
HRTF_ORIGINAL((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)*...
    length(DatosValidacion.SUJETO),:)=mat_test;
HRTF_APROX((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)*...
    length(DatosValidacion.SUJETO),:)=mat_recon;
SD((j-1)*length(DatosValidacion.SUJETO)+1:((j-1)+1)*...
    length(DatosValidacion.SUJETO),:)   =SD_k_fold;

end
DATOS_K_FOLD_RESULTADOS=table(SUJETOS,FOLD,COMPONENTES_PRINCIPALES,AZIMUT,...
    ELEVACION,EAR,HRTF_ORIGINAL,HRTF_APROX,SD);

save('DATOS_K_FOLD_RESULTADOS.mat','DATOS_K_FOLD_RESULTADOS',...
    '-v7.3','-nocompression')
```

ANEXO κ:ARGUMENTOS DE HRTFs

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