



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
ELÉCTRICA – INSTRUMENTACIÓN

FOTOANTROPOMETRÍA AUTOMATIZADA PARA PERSONALIZAR FUNCIONES
DE TRANSFERENCIA ACÚSTICA REFERIDAS A LA CABEZA (HRTF).

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN INGENIERÍA

PRESENTA:
EDGAR AUGUSTO TORRES GALLEGOS

TUTOR PRINCIPAL
DR. FELIPE ORDUÑA BUSTAMANTE, CCADET

COMITÉ TUTOR
DR. FELIPE ORDUÑA BUSTAMANTE, CCADET
DR. FERNANDO ARÁMBULA COSÍO
DR. RAFAEL DE LA GUARDIA GONZÁLEZ

MÉXICO, D. F. DICIEMBRE 2014

JURADO ASIGNADO:

Presidente: Dr. Ricardo Ruiz Boullosa
Secretario: Dr. Fernando Arámbula Cosío
Vocal: Dr. Felipe Orduña Bustamante
1er. Suplente: Dr. Boris Escalante Ramírez
2do. Suplente: Dr. Rafael De La Guardia González

Lugar donde se realizó la tesis:

Centro de Ciencias Aplicadas y Desarrollo Tecnológico,
Universidad Nacional Autónoma de México

TUTOR DE TESIS:

Dr. Felipe Orduña Bustamante



FIRMA

Este trabajo no habría sido posible si el apoyo de:

Coordinación de Estudios de Posgrado, Universidad Nacional Autónoma de México
Intel Tecnología de México, S.A. de C.V.
Beca Interna CCADET
Proyecto PAPIIT: “Ondas acústicas lineales y no lineales en guías de ondas”

Mi tutor:

Dr. Felipe Orduña Bustamante

Mi comité tutor:

Dr. Felipe Orduña Bustamante
Dr. Fernando Arámbula Cosío
Dr. Rafael de la Guardia González

Dr. Héctor Cordourier Maruri

Mis sinodales:

Dr. Felipe Orduña Bustamante
Dr. Fernando Arámbula Cosío
Dr. Rafael de la Guardia González
Dr. Ricardo Ruiz Boullosa
Dr. Boris Escalante Ramírez

Y todos mis amigos de

CCADET, UNAM
Intel Tecnología de México, S.A. de C.V.

A todos ellos, muchas gracias.

Índice general

Resumen	I
Congresos y publicación	II
1. Introducción	1
1.1. Medición directa de las HRTF	2
1.2. Objetivo	2
1.3. Sistema de personalización de HRTF	3
2. Antecedentes	5
2.1. Antropometría	6
2.1.1. Antropometría de la oreja	6
2.2. Visión computacional	7
2.2.1. Detección de rostros	7
2.2.2. Reconocimiento de rostros	8
2.2.3. Detección y reconocimiento de orejas	8
2.2.4. Caracterización de personas	8
2.3. Sobre la naturaleza de las HRTF	9
2.3.1. Pinna Related Transfer Function	10
2.4. Bibliotecas de HRTF	10
2.5. Personalización de HRTF	11
2.5.1. Personalización mediante la antropometría	12
2.5.2. Relación entre la antropometría y las HRTF	13
2.6. Modelado de HRTF a partir de imágenes	13
3. Fotoantropometría automatizada	14
3.1. Método	14
3.1.1. Sesión fotográfica	14
3.1.2. Modelos de forma activa	15

3.1.3. Entrenamiento del ASM	16
3.1.4. Ajuste del modelo	18
3.1.5. Antropometría a partir del modelo ajustado	21
3.2. Validación	23
3.2.1. Validación del ASM	24
3.2.2. Validación de la antropometría	24
4. Métodos de personalización de HRTF	27
4.1. Mejor coincidencia antropométrica	27
4.2. Regresión lineal múltiple	30
4.3. Síntesis con red neuronal artificial	32
5. Distorsión de espectro complejo	35
6. Validación de los métodos	38
6.1. Mejor coincidencia antropométrica	38
6.2. Comparación de los tres métodos	39
6.3. HRTF promedio y genérica	42
6.4. Efectos de los errores antropométricos en las HRTF	44
7. Personalización de HRTF a partir de la antropometría	47
8. Relevancia de los parámetros antropométricos	50
8.1. Método	51
8.2. Resultados	55
8.3. Comparación con la bibliografía	63
9. Discusión	66
9.1. Contribuciones originales	67
9.2. Recomendaciones	68
9.3. Trabajo futuro	68
10. Conclusiones	70
Bibliografía	72

Resumen

SE presenta un método automatizado para personalizar funciones de transferencia acústica referidas a la cabeza (HRTF) a partir de la información antropométrica de un sujeto, que a su vez se obtiene a partir de retratos digitales del sujeto y un algoritmo de visión computacional conocido como modelos de forma activa (ASM) [Cootes et al. *Comp. Vision and Image Understanding*, 61(1):38–59, 1995].

El trabajo presentado sigue la propuesta de [Zotkin et al. *WASPAA 2003*, pp. 157–160], y tiene ventajas sobre la solución de [Dellepiane et al. *Comp. Graph. Forum*, 27(7):1719–1727, 2008], pues en este trabajo se utilizan algoritmos que convergen más rápido, tanto el algoritmo de visión computacional, como el método de personalización de HRTF.

El método presentado es automático y usa un modelo en 2D más simple y eficiente en la representación del sujeto. El ASM se entrena antes de su aplicación, usando un grupo de imágenes anotadas manualmente; una vez entrenado, se usa para reconocer automáticamente los rasgos de la oreja, cabeza y torso usando las fotografías de un sujeto dado.

A partir del modelo se obtienen los parámetros antropométricos del sujeto en unidades métricas, usando el modelo ajustado sobre la fotografía, y en función de los ajustes de la cámara, como el campo de visión y la distancia al objetivo.

La información antropométrica se usa finalmente para personalizar las HRTF del sujeto por inferencia a partir de una base de datos que contiene la información antropométrica y las HRTF de un grupo de sujetos [Algazi et al. *WASPAA 2001*, pp. 99–102], idealmente representativo.

Los resultados del método son analizados y se concluye que la personalización de HRTF puede lograrse por medio de la fotoantropometría automatizada y su inferencia a partir de la base de datos usando el método de búsqueda de la mejor coincidencia antropométrica, que resultó el método preferible tras compararlo con otros métodos de personalización utilizando la métrica de la distorsión espectral y una nueva métrica propuesta que toma en cuenta la fase del espectro complejo.

Sin embargo, se presenta evidencia de que la base de datos de HRTF es aún insuficiente para proveer de un resultado preciso, lo que apunta a la necesidad de investigar los temas involucrados con mayor profundidad.

Congresos y publicación

El trabajo presentado en esta tesis fue presentado, está en proceso de publicación o será presentado en las siguientes fuentes:

- E. A. Torres Gallegos, F. Orduña Bustamante, y F. Arámbula Cosío. Fotoantropometría automatizada para personalizar funciones de transferencia acústica referidas a la cabeza (HRTF). En SOMI XXVII, Culiacán, Sinaloa, Octubre de 2012.
- E. A. Torres-Gallegos, F. Orduña-Bustamante, F. Arámbula-Cosío, System for automatic personalization of head-related transfer functions (HRTF) based on computer vision photo-anthropometry and inference from a database. 166th Meeting of the Acoustical Society of America, San Francisco, California, Diciembre de 2013.
- E. A. Torres Gallegos, F. Orduña Bustamante, y F. Arámbula Cosío. Sistema de foto-anthropometría automatizada y su uso en la personalización de funciones de transferencia acústica referidas a la cabeza (HRTF). En SOMI XXIX, Puerto Vallarta, Jalisco, Octubre de 2014.
- E. A. Torres-Gallegos, F. Orduña-Bustamante, F. Arámbula-Cosío, “Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database” (en proceso de corrección), Applied Acoustics.

Capítulo 1

Introducción

EL sistema auditivo humano es capaz de discernir la dirección de procedencia del sonido que incide en los oídos; esto es posible gracias a que la información que necesita el cerebro para resolver la dirección de procedencia se encuentra mezclada con el sonido que incide en cada oído, a este fenómeno se le conoce como *sonido biaural*, y a la información que transmite la ubicación espacial de la fuente se le conoce como *función de transferencia acústica relacionada con la cabeza* o HRTF¹ [1–4].

Se puede entender la naturaleza de las HRTF como filtros acústicos, y son producto de las interacciones (difracción, reflexión, atenuación, absorción) del frente de onda con la cabeza, cuello, torso y orejas de una persona. Estas interacciones dependen completamente tanto de la dirección de procedencia del sonido, así como de la forma del sujeto (tamaño de la cabeza, pliegues de la oreja, etc.), la cual puede caracterizarse con técnicas de antropometría; y debido a que todas las personas son diferentes, las HRTF son únicas para cada persona; también son diferentes entre cada oreja de un mismo sujeto, y para cada ángulo de procedencia.

Es posible sintetizar muestras de sonido biaural en lo que se conoce como *tecnología biaural* [5] para crear entornos acústicos virtuales. Esto es posible mediante la convolución de dos componentes: la primera es una muestra de audio convencional (monoaural o estereofónico, a la que se le añadirá el efecto espacial) y la segunda son las HRTF de una persona (una para cada oído), específicas de la dirección de procedencia que se desea recrear.

Las aplicaciones de la tecnología biaural, y los entornos acústicos virtuales, son variadas, desde audio para personas invidentes, realidad virtual, videojuegos, etcétera [6]; en telecomunicaciones es útil por su capacidad de mejorar la intelli-

¹HRTF, del inglés *Head Related Transfer Function*.

bilidad de un mensaje hablado, especialmente en condiciones adversas con niveles altos de ruido y reverberación [7–10].

Actualmente en la mayoría de las aplicaciones de sonido binaural se utilizan HRTF genéricas, grabadas sobre un maniquí, cuyas características antropométricas corresponden al promedio de una población [11, 12]. Esto es un inconveniente pues el uso de HRTF genéricas introduce errores perceptuales [13–15] debido a la individualidad de las HRTF. Estos errores se previenen si se utilizan las HRTF personales de cada usuario, las cuales se obtienen, principalmente, de la medición directa sobre el sujeto.

1.1. Medición directa de las HRTF

Es posible medir las HRTF de cada persona, para esto es necesario colocar un micrófono en la entrada o el interior del conducto auditivo de cada oreja; también es necesario colocar una fuente sonora² en una posición conocida alrededor del sujeto. La HRTF se obtiene como la relación entre los espectros del sonido captado en el micrófono y el emitido por la fuente.

Esta medición directa de las HRTF sobre un sujeto es un procedimiento complejo y tardado; se requiere la disponibilidad de un recinto aislado del ruido, idealmente una cámara anecoica [16], equipo especializado y calibrado, además de micrófonos específicos para este tipo de medición, y que podrían causar molestias al sujeto; aunado a esto, se requiere que el sujeto permanezca inmóvil durante el transcurso de las mediciones, las cuales deben repetirse para cada ángulo de excitación. Algunos trabajos proponen métodos más rápidos de medición, sin embargo, estos requieren arreglos experimentales más complejos y costosos [17, 18].

Con el estado actual de estas técnicas, la medición de las HRTF de cada sujeto para su uso en aplicaciones de tecnología binaural es un procedimiento prohibitivo que impide el uso generalizado de la tecnología binaural. Se requiere entonces un método más rápido y accesible, de uso generalizable, para personalizar las HRTF de cada sujeto.

1.2. Objetivo

El contenido de este trabajo gira en torno al siguiente objetivo:

Obtener una colección de HRTF específicas para cada persona, seleccionadas o sintetizadas en forma automática a partir de los retratos fotográficos del sujeto, y evitando la medición directa de las HRTF.

²Para este propósito se puede usar un barrido en frecuencia, ruido blanco, secuencias MLS, impulsos, etc.

1.3. Sistema de personalización de HRTF

En este trabajo se propone y desarrolla un sistema para personalizar HRTF a partir de la información antropométrica de un sujeto. Esta información se obtiene a su vez a partir de retratos del mismo mediante técnicas de visión computacional. El sistema está compuesto por tres módulos principales, cada uno con la siguiente función: (1) identificación automática de rasgos antropométricos utilizando algoritmos de visión computacional, (2) fotoantropometría automática, y (3) inferencia de las HRTF a partir de la antropometría.

En el primer paso, una colección de retratos del sujeto es procesada con un algoritmo de visión computacional con el que se obtiene un *modelo de distribución de puntos* (PDM)³ entre los que se encuentran los puntos *somatométricos* que identifican a los rasgos antropométricos clave.

En el siguiente paso se mide en forma automática la distancia entre los puntos somatométricos y éstas se convierten a unidades métricas, esto corresponden a la antropometría del sujeto. Los pasos (1) y (2) se realizan en forma automática con software propietario, avanzando el estado del arte propuesto en [19].

Finalmente, en el último paso se usa la información antropométrica para inferir las HRTF del sujeto; para esto es posible utilizar varios métodos, desde simulación numérica, hasta su selección directa a partir de una base de datos de HRTF que, idealmente, sea lo suficientemente representativa para personalizar las HRTF de toda una población.

En este trabajo se comparan tres métodos de personalización de HRTF, los cuales utilizan una base de datos: (a) selección directa de la HRTF que corresponda al sujeto con la mayor semejanza antropométrica dentro de la base de datos; (b) síntesis de HRTF usando una *regresión lineal múltiple* (MLR⁴) entre la antropometría y una base ortogonal obtenida utilizando el análisis de componentes principales; y (c) síntesis de la HRTF usando una *red neuronal artificial* (ANN⁵), entrenada con la base de datos con la antropometría como información de entrada y la HRTF como salida.

Este sistema propuesto para personalización de HRTF es conveniente; a diferencia de la medición directa, en la que se requiere la participación y paciencia del sujeto por varias horas, mientras que en el sistema propuesto solamente se requiere de su participación por unos cuantos minutos, necesarios para obtener sus retratos en una sesión fotográfica.

A largo plazo, el objetivo de futuras iteraciones de este trabajo es integrar el sistema en un dispositivo portátil de entretenimiento o comunicación (ej. teléfono inteligente), con el que será posible obtener las HRTF del usuario tras pedirle al

³PDM, del inglés *Point Distribution Model*.

⁴MLR, del inglés *Multiple Linear Regression*.

⁵ANN, del inglés *Artificial Neural Network*.

sujeto que se retrate con la cámara del dispositivo.

Contenido del trabajo

El resto de este trabajo está organizado en los siguientes capítulos: En el capítulo 2 se presenta el estado del arte en las áreas de visión computacional, fotoantropometría y personalización de HRTF.

En el capítulo 3 se resuelve el problemas de la fotoantropometría automatizada, se presenta el algoritmo empleado y el procedimiento de validación. Una vez que se tiene la información antropométrica se aborda el segundo objetivo de este trabajo: la personalización de las HRTF. En el capítulo 4 se proponen tres métodos de personalización de HRTF; en el capítulo 5 se propone una nueva métrica para evaluar los resultados de la personalización. Con todo esto es posible realizar la personalización de las HRTF de varios sujetos; esto se presenta en el capítulo 7, y los métodos de personalización de HRTF se evalúan en el capítulo 6.

El estudio concluye con un análisis de la relevancia que tienen los parámetros antropométricos en la personalización de las HRTF, esto se presenta en el capítulo 8. Finalmente en los capítulos 9 y 10 se discuten los resultados y se presentan las conclusiones de esta investigación.

Capítulo 2

Antecedentes

El objetivo de este trabajo se sustenta principalmente en los siguientes antecedentes directos:

1. Un grupo de investigadores del laboratorio de CIPIC (de la Universidad de California en Davis), dirigidos por Algazi, publicaron en 2001 una base de datos de HRTF de 45 sujetos con su correspondiente información antropométrica [20].

2. Zotkin [19, 21] propuso el uso de esta información para seleccionar HRTF compatibles con un nuevo sujeto, siguiendo la hipótesis de que dos personas con rasgos antropométricos similares tienen HRTF compatibles. Este paso simplifica el problema de la personalización de HRTF a uno de comparación antropométrica. Para resolver este nuevo problema usaron la técnica de fotoantropometría manual (no automática), mediante un sistema de software propietario.

3. Nishino [22] propuso el uso del análisis de componentes principales (PCA) [23, 24] para descomponer su base de datos de HRTF en una base ortogonal con la cual se pueden sintetizar nuevas HRTF al variar las ponderaciones de las *componentes principales* y mezclándolas mediante una combinación lineal. Para calcular la ponderación que debe asignar a cada componente principal realizó una regresión lineal múltiple (MLR) entre la antropometría y las ponderaciones de las componentes principales. Así, para obtener las HRTF de un nuevo sujeto, basta con multiplicar su antropometría con la matriz de regresión para calcular las nuevas ponderaciones de las componentes principales. Para validar su trabajo, Nishino propone el uso de la métrica de distorsión espectral (SD), que mide el error entre dos HRTF.

4. Este trabajo no constituye el primer esfuerzo en personalizar las HRTF a partir de fotografías; Dellepiane [25] presentó un método de ajuste de modelos genéricos tridimensionales, de la cabeza y la oreja, sobre fotografías del sujeto. Una vez ajustado, usa el modelo para generar las HRTF con el método de elemento

de frontera (BEM). Esta idea fue propuesta originalmente por Duraiswami [26]. Sin embargo este método tiene desventajas importantes; requiere de varios días de cómputo para obtener las HRTF. Opuesto a esto, el método propuesto en este trabajo utiliza modelos bidimensionales, más sencillos (varios órdenes de magnitud), con lo que se llega a una solución en cuestión de segundos.

Se presenta a continuación una revisión más detallada del estado del arte en el campo fotoantropometría automatizada, y personalización de HRTF.

2.1. Antropometría

La *antropometría* es la medición de los rasgos anatómicos de una persona, y el término se refiere tanto al proceso de medición como a los datos que se obtiene a través de éste. La antropometría es útil en investigación médica, social y en otras ciencias. En medicina es útil porque permite obtener una estadística de las dimensiones de las poblaciones humanas y correlacionar las desviaciones con patologías genéticas. Por ejemplo, el síndrome de Down puede detectarse realizando la antropometría a un feto o un neonato [27], o puede emplearse para caracterizar e identificar personas [28], etc.

La antropometría es útil para propósitos de investigación acústica en la elaboración de maniqués de medición a partir de la antropometría promedio de una población, como por ejemplo: [29, 30].

La antropometría normalmente se realiza directamente sobre el cuerpo del sujeto, utilizando diversos instrumentos de medición. En contraste con este punto, la *foto-antropometría* es la técnica de medición sobre una fotografía del sujeto [27]. En el pasado se realizaba sobre una impresión fotográfica; actualmente se hace en la computadora, sobre una fotografía digital. En cualquier caso, era necesario que un especialista humano realice la medición.

Finalmente, la *foto-antropometría automatizada* es un campo de investigación actual que se dedica al desarrollo de técnicas y algoritmos computacionales para obtener la antropometría de una persona en forma automatizada.

2.1.1. Antropometría de la oreja

La forma de la oreja, caracterizada por medio de su antropometría, es un tema bien estudiado [31–37], principalmente con propósitos médicos, estéticos y forenses. Se sabe que la oreja, especialmente el lóbulo, continúa su crecimiento a lo largo de la vida del sujeto, este, sin embargo, no es un factor relevante para este trabajo, pues la personalización se puede repetir para ajustarse a cualquier cambio de antropometría con el transcurso del tiempo. También es claro que la forma

de ambas orejas no es del todo simétrica, hay diferencias de forma que pueden representar diferencias de hasta 10 % en la antropometría (excluyendo casos de malformaciones congénitas). Esto tampoco representa una complicación, pues en este trabajo el proceso de antropometría se puede realizar indiferentemente para ambas orejas.

2.2. Visión computacional

Se entiende por visión computacional al grupo de técnicas digitales para adquirir, procesar, analizar y entender información compleja multidimensional de nuestro entorno para propósitos de exploración científica y técnica [38, 39].

En esta investigación se requiere aplicar estas técnicas para realizar el proceso de fotoantropometría en forma automática. El problema de visión computacional suele ser dividido en dos partes: la detección y el reconocimiento. La detección es la etapa en la cual, tras alimentar al sistema de visión con una imagen digital, el programa busca la presencia de un objeto. La siguiente etapa es el reconocimiento; es aquí donde se ubican los rasgos característicos del objeto (en este caso los puntos somatométricos de una persona) y se realizan las mediciones pertinentes.

La naturaleza del problema, es decir, la fotoantropometría, requiere otra división del proceso de visión computacional, por un lado se tiene la necesidad de detectar y reconocer el rostro, cabeza y torso del sujeto; y por el otro se tiene la detección y el reconocimiento de los rasgos de la oreja. Son dos problemas complejos.

2.2.1. Detección de rostros

El rostro humano es un objeto dinámico con un alto grado de variabilidad en su apariencia, para resolver el problema de la detección se han propuesto varios algoritmos [40] que se concentran en la detección del rostro en imágenes con perspectiva frontal [41]; el caso de la perspectiva lateral es menos estudiado.

Los algoritmos de detección de rostros se clasifican en:

- Algoritmos basados en las características: a) Análisis de bajo nivel: orillas, niveles de gris, color, movimiento y medición generalizada. b) Análisis de características: búsqueda de características y análisis de constelación. c) Modelos de forma activa: serpientes, moldes deformables y modelos de distribución de puntos.
 - Algoritmos basados en la imagen: a) Métodos de subespacios lineales. b) Redes neuronales. c) Algoritmos estadísticos
-

2.2.2. Reconocimiento de rostros

Durante los últimos veinte años, la investigación se ha enfocado en automatizar completamente los sistemas de reconocimiento de rostros, separando el problema en la detección del rostro en una imagen dada y en la extracción de rasgos como los ojos, boca, etc. Sin embargo, las técnicas de extracción de rasgos no son aún lo suficientemente confiables [42]. Por ejemplo, muchas técnicas de ubicación del ojo no funcionan cuando el ojo está cerrado.

Las técnicas de reconocimiento facial (FRT)¹ se clasifican en dos grupos principales, dependiendo de si analizan fotografías o video del sujeto [43]. El problema generalmente se formula como el reconocimiento de objetos tridimensionales (3D) a partir de imágenes bidimensionales (2D) [44]. Anteriormente se trataba como un problema de reconocimiento de patrones.

2.2.3. Detección y reconocimiento de orejas

En lo que respecta a la antropometría de la oreja, el problema también divide en detección [45] y reconocimiento, y se comenzó a estudiar hace tan sólo una década. El problema del reconocimiento ha sido abordado con diferentes algoritmos entre los cuales se tiene: Redes neuronales [46], búsqueda genética sobre imágenes de intensidad [47], reconocimiento 3D usando descripciones locales de formas de superficie [48], ICP² de dos pasos [52], extracción de parámetros utilizando transformación de campos de fuerza [53, 54], entre otros [45, 48, 52, 55–62].

Originalmente, los algoritmos de reconocimiento de orejas eran muy sensibles al ruido que introducían la presencia de aretes y cabello cubriendo parcialmente la zona; los algoritmos recientes hacen coincidir un modelo tridimensional de la oreja en la imagen, y han alcanzado un nivel de éxito en el reconocimiento de 97.6 % [63].

2.2.4. Caracterización de personas

Las técnicas de visión computacional ya han sido empleadas para caracterizar personas, por ejemplo, para obtener los puntos somatométricos a partir de modelos 3D de todo el cuerpo [64], para obtener la antropometría de la cabeza [65, 66], para caracterizar orejas [67–70], o la cara [71].

Un primer esfuerzo para obtener la antropometría del sujeto mediante técnicas de visión computacional [72], y con el propósito de personalizar HRTF, extrae las

¹Del inglés *Face Recognition Techniques*.

²El ICP, del inglés *Interactive Closest Point* es un algoritmo para hacer coincidir formas tridimensionales [49–51].

medidas del cuello, cabeza y torso para aplicarlas a un modelo sencillo donde la cabeza y el torso se aproximan con esferoides.

2.3. Sobre la naturaleza de las HRTF

En la audición humana, debido a la naturaleza de propagación del sonido y a la diferencia en la posición de las orejas (diametralmente opuestas), el sonido que registra un oído es generalmente distinto al que registra el otro. A este efecto se le conoce como sonido biaural y permite a las personas ubicar una fuente sonora en el espacio a su alrededor [73].

La teoría Dúplex, propuesta por Lord Rayleigh en 1907 [74] y estudiada en otros trabajos [75–77], propone dos efectos para explicar la capacidad de ubicar una fuente en el espacio; uno es la diferencia de intensidad interaural (ILD) y el otro la diferencia de tiempo interaural (ITD).³ En la actualidad se sabe que esta teoría funciona bien para sonidos en el intervalo de bajas frecuencias [78]; para el caso de las altas frecuencias, con longitudes de onda del orden de centímetros y debido a la interacción del sonido con la oreja, la cabeza y el torso, el problema es más complejo [15, 20], se recurre entonces al concepto de función de transferencia acústica relacionada con la cabeza (HRTF)⁴, que puede entenderse como un filtro acústico que modifica al sonido, y cuyos rasgos espectrales permiten al cerebro interpretar la dirección de procedencia del sonido.

Para propósitos de localización, los humanos son sensibles principalmente a la magnitud de la HRTF, no la fase [79]. Esto sugiere que en trabajos de personalización de HRTF debe ponerse mayor atención al error de magnitud. Los rasgos espectrales de las HRTF, es decir, los picos y valles del espectro, son las características significativas que son perceptualmente importantes para localizar la fuente en el espacio [80–82]. Las HRTF son diferentes entre cada oído y entre cada ángulo de procedencia, además de esto, las HRTF también varían en función de la distancia a la fuente [83, 84].

También se sabe que el cerebro es capaz de ajustarse a cambios en la forma del oído, es decir, puede volver a aprender una HRTF [85] usando pistas visuales de la localización del sonido. Esto podría funcionar como una nueva forma de “personalización”; usar HRTF genéricas, pero ajustar a la persona a estas HRTF genéricas mediante pistas visuales. Esta idea es congruente con los resultados presentados en [86]. Sin embargo, también se sabe que la visión no es necesaria para la localización de una fuente sonora [87].

³Del inglés *Interaural Level Difference* e *Interaural Time Difference*.

⁴Del inglés *Head-Related Transfer Function*.

2.3.1. Pinna Related Transfer Function

Algunos autores simplifican el problema de la personalización de las HRTF y se enfocan solamente en la interacción del sonido con la pinna, lo que resulta en las *funciones de transferencia relacionadas con la pinna* o PRTF⁵. Entre ellos tenemos los trabajos [88–94]. No debe olvidarse que estos no constituyen una solución completa al problema de personalización de HRTF.

Sin embargo la información obtenida en estos estudios no deja de ser útil, por ejemplo es posible analizar la relación entre los pliegues de la pinna y las componentes espectrales de las PRTF [82, 88, 93–97]; esto constituye un gran paso en la investigación.

2.4. Bibliotecas de HRTF

Recopilar una base de datos de HRTF es una actividad ardua; varios grupos se han dado a la tarea y han publicado sus bibliotecas para propósitos de investigación; otros grupos se reservan sus datos [1, 12, 20, 22, 98–110]. Las más relevantes son:

- (1995) MIT-Media-Lab KEMAR HRTF Database [12] grabada sobre un maniquí KEMAR.
- (2001) CIPIC HRTF database [20] grabada sobre 43 sujetos, incluye información antropométrica de 27 parámetros. Agregan las HRTF del maniquí KEMAR con dos pinnas diferentes. Es la más empleada en la literatura.
- (2003) LISTEN project, del laboratorio IRCAM y en colaboración con AKG, incluye las HRTF de 51 sujetos [100].
- (2007) CIAIR [22], incluye 86 sujetos, con información de 10 parámetros antropométricos.

Debido a la complejidad para un sólo grupo de investigación de armar una biblioteca con una colección grande de sujetos, se han realizado actividades para consolidar las distintas bases de datos de HRTF en una base de datos única [111, 112], con el mayor número de sujetos disponibles.

El proceso de medición de las HRTF también ha sido objeto de estudio, especialmente la repetibilidad de la medición [98, 106, 113]. También se han hecho esfuerzos por simplificarla [114].

⁵PRTF, siglas en inglés de *Pinna Related Transfer Function*.

Reducción de las dimensiones de la biblioteca Una HRTF puede caracterizarse de acuerdo a las coordenadas que ocupa en un espacio multidimensional, donde cada coordenada es la magnitud de cada una de las componentes de frecuencia. Personalizar una HRTF implica predecir las coordenadas de todas las componentes de frecuencia. Ante la complejidad de este problema, se intenta “reducir la dimensionalidad” [115, 116] de las HRTF, transfiriéndolas a un espacio de menos dimensiones, con pérdida mínima de información, simplificando así el problema de la personalización.

La reducción dimensional y compresión de la base de datos se ha realizado con diversas técnicas como: análisis de componentes principales (PCA⁶) [24, 117–119], onduletas (*wavelets*) [120–124], LLE⁷ [125–127], Hankel Total Least Squares Method [128, 129], entre otros [130–134].

Interpolación de HRTF La interpolación es útil para construir una colección de HRTF de un solo sujeto a partir de pocas mediciones, reduciendo el tiempo dedicado a la medición de las HRTF en el laboratorio [126, 135–154]. También se ha propuesto el uso de redes neuronales artificiales para interpolación de HRTF [155–158]. Es importante aclarar que el concepto de interpolación de HRTF no está relacionado con el proceso de personalización de HRTF.

2.5. Personalización de HRTF

Opuesto al uso de HRTF genéricas [13], en la literatura se han presentado varios métodos para personalizar HRTF; estos se pueden clasificar en las siguientes categorías:

Métodos subjetivos Algunos autores han propuesto delegar la tarea de la personalización al sujeto que empleará las HRTF, permitiéndole por medio de una prueba auditiva que sea éste quien decida qué HRTF de entre una biblioteca es la que le confiere la sensación espacial más acertada, entre estos métodos tenemos la selección [159], síntesis [160, 161], y ecualización [162].

Agrupamiento de HRTF Si se considera que las HRTF ocupan un espacio multidimensional en el que cada dimensión es alguna de sus características, es posible entonces calcular la cercanía de las HRTF en este espacio con la finalidad de agrupar aquellas que se encuentren más próximas a otras; la hipótesis principal consiste en que dos HRTF cercanas son intercambiables entre los sujetos correspondientes [163–166].

⁶PCA, siglas de *Principal Components Analysis*.

⁷LLE, siglas de *Locally Linear Embedding*.

Escalamiento de las HRTF También se ha propuesto que una HRTF grabada sobre un sujeto de antropometría conocida, sea asignada a otro sujeto mediante un escalamiento en frecuencia para compensar por los cambios de antropometría, principalmente en lo que respecta al tamaño de la *pinna* y el ancho de la cabeza [167–169]

Modelado de las HRTF Otros investigadores han enfrentado el problema de la personalización mediante distintas técnicas de modelado, entre ellas tenemos: (a) modelo sencillo de la cabeza y el torso, aproximando estos últimos como esferoides [78, 170–178], (b) suma de elementos (ecos) de la *pinna*, la habitación, el hombro y retrasos y sombra de la cabeza [179, 180], (c) modelado por medio de filtros paramétricos [181, 182], y (d) otros [183, 184], entre los que destacan algunos que utilizan la antropometría para modificar el modelo [89, 185, 186].

Simulación numérica Las HRTF pueden calcularse con gran precisión usando algoritmos de simulación numérica, sin embargo tienen el inconveniente de un gran costo computacional, requiriendo hasta varios días para obtener un resultado. Entre los métodos usados tenemos: elemento de frontera (BEM⁸) [83, 187–195], diferencias finitas [196] y *boundary sound representation* [197].

2.5.1. Personalización mediante la antropometría

Reconociendo la relación que existe entre la morfología de un sujeto, caracterizado mediante su antropometría, y las HRTF, múltiples autores han propuesto métodos para seleccionar o sintetizar HRTF partiendo de la información antropométrica, entre ellos tenemos:

Selección de HRTF, minimizando la distancia antropométrica. Una de las primeras soluciones propuestas al problema de la personalización de HRTF a partir de la antropometría del sujeto fue la selección, dentro de la base de datos, del sujeto más parecido en términos antropométricos [19, 198], buscando al sujeto de la base de datos para el cual la distancia o error entre antropometrías sea mínima.

Síntesis de HRTF mediante componentes principales La descomposición de las HRTF en una base ortogonal por medio del análisis de componentes principales (PCA) [24] resulta en una base, con esta se pueden formar HRTF nuevas para cualquier sujeto [117, 118, 161, 199–201], siempre y cuando la base de datos utilizada sea lo suficientemente representativa.

⁸BEM, siglas del inglés *Boundary Element Method*.

La base permite la síntesis de HRTF mediante la mezcla de las componentes principales, asignando a cada componente una ponderación que se obtiene a partir de regresiones entre las componentes principales y la antropometría. Utilizando este método de regresión se han publicado los siguientes trabajos: [22, 202–212].

Redes neuronales Las redes neuronales artificiales han sido empleadas para reemplazar la regresión [213], sin embargo, este método es lineal, desaprovechando la capacidad de las redes neuronales de ajustar funciones no lineales.

2.5.2. Relación entre la antropometría y las HRTF

Para lograr exitosamente la personalización de las HRTF, es necesario entender la relación entre morfología de la personas (caracterizada por medio de su antropometría) y los efectos que ésta produce en la formación de las HRTF. Un mejor entendimiento de estas relaciones permitirá en el futuro desarrollar métodos de personalización de HRTF más eficaces. En este trabajo se analiza y discute este punto, y se presentan nuevos resultados que se agregan a los de otros autores [214–216]. Se trata, sin embargo, de una línea de investigación inconclusa.

En el estado actual de la investigación, los resultados que son útiles son aquellos en los que se estudia, mediante métodos estadísticos o selección arbitraria [217, 218], la relevancia de los parámetros antropométricos que conforman la base de datos de CIPIC.

2.6. Modelado de HRTF a partir de imágenes

Siguiendo una idea similar a la que se presenta en este trabajo, es decir, obtener información del sujeto a partir de sus retratos y con esta información personalizar las HRTF, dos grupos han presentado trabajos: uno obtiene información de los pliegues de la *pinna* para relacionarlo con los rasgos espectrales de la HRTF [219, 220], y otro se encarga del ajuste de modelos tridimensionales a la fotografía y obtención de las HRTF utilizando el método de elemento de frontera [25] los inconvenientes de este método (gran costo computacional) fueron mencionado anteriormente. La investigación en esta dirección sigue en curso.

Capítulo 3

Fotoantropometría automatizada

3.1. Método

En esta sección se propone el uso de un algoritmo de visión computacional con el propósito de realizar la foto-antropometría automatizada de un sujeto cualquiera. El método presentado se enfoca exclusivamente en obtener los rasgos (o parámetros) antropométricos que serán empleados en este trabajo para personalizar las HRTF, sin embargo, el método puede alterarse para ajustarse, mediante un nuevo entrenamiento, a aplicaciones futuras de antropometría o caracterización de distintos objetos a partir de fotografías.

El método consiste en cuatro etapas: 1) sesión fotográfica para obtener retratos de varios sujetos, 2) entrenamiento de un algoritmo de visión computacional para generar un modelo deformable de los contornos de los sujetos, 3) uso del algoritmo de visión para obtener un modelo ajustado a los contornos de un nuevo sujeto, y 4) obtención de la antropometría del nuevo sujeto a partir de su modelo ajustado.

3.1.1. Sesión fotográfica

Los retratos de los sujetos se obtuvieron en un estudio fotográfico bajo condiciones controladas de iluminación, fondo, posición, postura, distancia a la cámara, etc. Como estudio se utilizó el diaporama del CCADET, de fondo azul, lo que permite la posterior remoción del fondo en el procesamiento digital de los retratos.

Se utilizó una cámara DSLR Canos EOS 500D con dos lentes primas con distancias focales de 28 mm y 50 mm. El primer lente se utilizó para los retratos de frente, permitiendo encuadrar el ancho del torso de los sujetos. La otra lente se usó para fotografías de perfil, permitiendo obtener más detalle en la zona de la oreja.

La distancia del objetivo al plano focal se fijó en 1.5 m. Otros detalles, incluyendo el arreglo de la iluminación, se muestran en las figuras (3.1) y (3.2), y en la figura (3.3) se muestran ejemplos de los retratos obtenidos de uno de los sujetos, para el caso de frente, perfil y un aumento a la zona de la oreja.

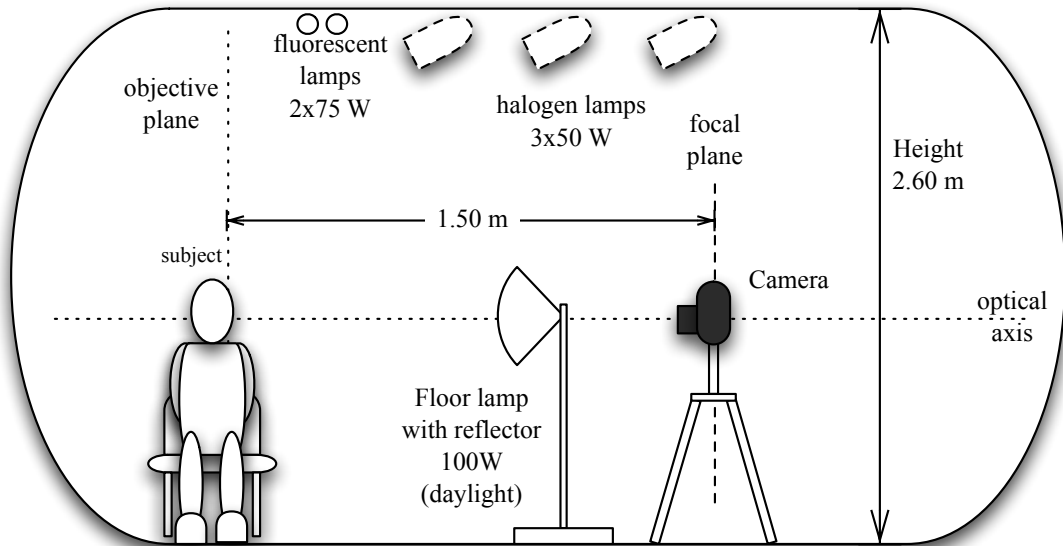


Figura 3.1: Vista lateral del arreglo fotográfico.

3.1.2. Modelos de forma activa

Los modelos de forma activa (ASM)¹ [221] son un algoritmo de visión computacional que consiste en el ajuste iterativo de curvas deformables a los bordes de interés en una imagen. Las deformaciones solamente pueden ocurrir en direcciones específicas, de modo que la forma fundamental del modelo nunca se pierde, esto impide que el modelo de una oreja se deforme hasta ajustarse a otro contorno, como el de una mano, por ejemplo.

El algoritmo ASM consiste de dos etapas: 1) entrenamiento manual del modelo, y 2) ajuste autónomo del modelo sobre la fotografía de interés.

Se diseñaron y entrenaron tres modelos independientes (estos se muestran en la primera columna de la figura (3.4) y se explican en la siguiente sección). El primer modelo ajusta a la vista frontal del sujeto, el segundo ajusta a la vista lateral y el tercero a la oreja del sujeto desde una perspectiva lateral.

Una vez entrenados, estos modelos son capaces de ajustarse, en forma autónoma, a los perfiles de sujetos nuevos cuyos retratos sean obtenidos en condiciones

¹Siglas del inglés *Active Shape Models*.

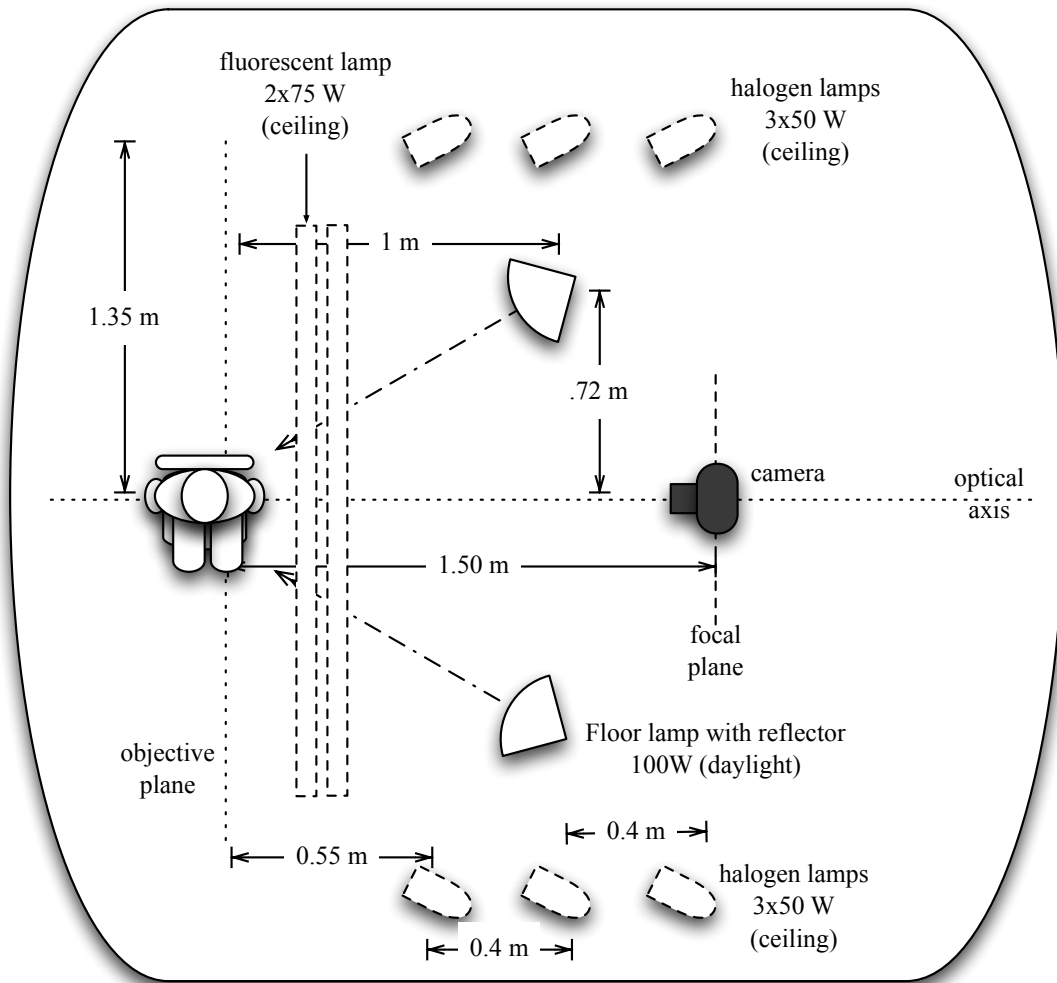


Figura 3.2: Vista superior del arreglo fotográfico.

similares. Es posible entonces obtener la antropometría del sujeto retratado si se mide la distancia entre puntos específicos del modelo ajustado, como se mostrará en la sección 3.1.5.

3.1.3. Entrenamiento del ASM

Los modelos consisten en una colección de puntos cuyas coordenadas deben anotarse manualmente sobre un grupo de imágenes de entrenamiento. El modelo de la oreja se entrenó usando nueve variaciones fotográficas (cada una desde una perspectiva diferente) de las dos orejas de once sujetos. Esto constituye un total de 198 formas distintas de orejas usadas para el entrenamiento. La capacidad del ASM para ajustarse a nuevos casos aumenta al usar una mayor cantidad y



Figura 3.3: Fotografías tomadas del sujeto 8: vista frontal, vista lateral y área de la oreja.

diversidad de formas de entrenamiento.

Para el caso de los modelos frontal y lateral, durante el entrenamiento se observó que convergían sin requerir una gran cantidad de casos de entrenamiento, por lo que fueron entrenados con sólo una y dos perspectivas de cada sujeto, respectivamente.

El proceso de entrenamiento, igual para los tres modelos, se explica con más detalle a continuación:

Sobre cada fotografía, p , se anota manualmente una colección de N puntos (que constituyen al modelo completo), cada uno con coordenadas (x_n, y_n) . Cada punto se almacena en forma secuencial dentro de un vector columna $\tilde{\mathbf{x}}_p = (x_1, y_1, \dots, x_N, y_N)^T$, donde T denota transposición.

El siguiente paso es obtener la forma promedio de todos los vectores de coordenadas, de la siguiente forma:

$$\bar{\mathbf{x}} = \frac{1}{P} \sum_{p=1}^P \tilde{\mathbf{x}}_p \quad (3.1)$$

donde P es el número total de imágenes anotadas. Este promedio se utilizará como la forma inicial del modelo durante la etapa de ajuste.

En la primera columna de la figura (3.4) se muestran, con círculos grises, los puntos que constituyen cada vector de entrenamiento, $\tilde{\mathbf{x}}_p$, y con círculos negros, la forma promedio, $\bar{\mathbf{x}}$.

Para mantener la claridad de las figuras, para cada caso se muestra sólo uno de cada tres puntos que constituyen a los modelos. El número total de puntos

anotados que se usaron en cada modelo fue de 189 puntos para el modelo de la oreja, 403 puntos para el modelo lateral y 761 puntos para el modelo frontal.

Para continuar con el entrenamiento, la forma promedio $\bar{\mathbf{x}}$ se sustrae de todos los vectores de entrenamiento, de esto resulta un grupo de vectores de desviaciones, es decir:

$$\tilde{\mathbf{x}}_p^d = \tilde{\mathbf{x}}_p - \bar{\mathbf{x}} \quad (3.2)$$

que se almacenan dentro de la matriz $\mathbf{X}_{2 \cdot N \times P}^d$. De estas desviaciones y utilizando el análisis de componentes principales (PCA) [24] se obtiene una base ortogonal que describe los ejes de deformación de los modelos. Esta base será fundamental para deformar el modelo durante la etapa de ajuste, permitiéndole tomar la forma de un sujeto nuevo con el que no fue entrenado; esto es posible con una simple variación de los coeficientes (ponderaciones) w_n de cada una de las componentes principales.

A modo de ejemplo, en las columnas (a), (b) y (c) de la figura (3.4) se puede apreciar visualmente la deformación de cada modelo al variar la ponderación, en la dirección positiva y negativa, de cada una de las tres primeras componentes.

En el último paso del entrenamiento, el fondo azul se remueve en forma automática aplicando un umbral que discrimina el tono azul; las imágenes de entrenamiento se convierten a escala de gris, y de éstas se extrae un grupo de muestras de pixeles a lo largo de las líneas normales al modelo, que atraviesan cada uno de los puntos de $\tilde{\mathbf{x}}_p$. Estas muestras de pixeles se almacenan en vectores $\tilde{\mathbf{g}}_n^p$, uno por cada punto n y fotografía p , y se conocen como *perfiles de gris*; contienen información sobre el entorno de cada punto del modelo, lo que resultará útil durante la etapa de ajuste, cuando sea necesario encontrar la mejor ubicación de cada punto sobre una imagen nueva.

Los perfiles de gris se almacenan en una matriz $\mathbf{G}_{L \times N}^p$, donde L es el número de pixeles de cada perfil. Al final del entrenamiento, es necesario conservar las componentes principales, el modelo promedio $\bar{\mathbf{x}}$ y los perfiles de gris $\mathbf{G}_{L \times N}^p$. Este procedimiento se realiza una sola vez, y el resultado se guarda para todos los ajustes posteriores.

3.1.4. Ajuste del modelo

Durante la etapa del ajuste del modelo a una fotografía nueva, el algoritmo ASM modifica la escala, posición, rotación y forma (ponderación de las componentes principales) del modelo, contenido en el vector \mathbf{X} , para hacerlo coincidir lo mejor posible a los perfiles de interés sobre la imagen.

El procedimiento de ajuste del modelo se describe a continuación. El modelo \mathbf{X} se inicializa con la forma promedio del entrenamiento, $\mathbf{X} = \bar{\mathbf{x}}$, y se coloca sobre la imagen en una posición arbitraria \mathbf{x}_c , rotación θ y escala s . A partir de este estado, cada uno de los puntos debe ser desplazado a una mejor posición, de

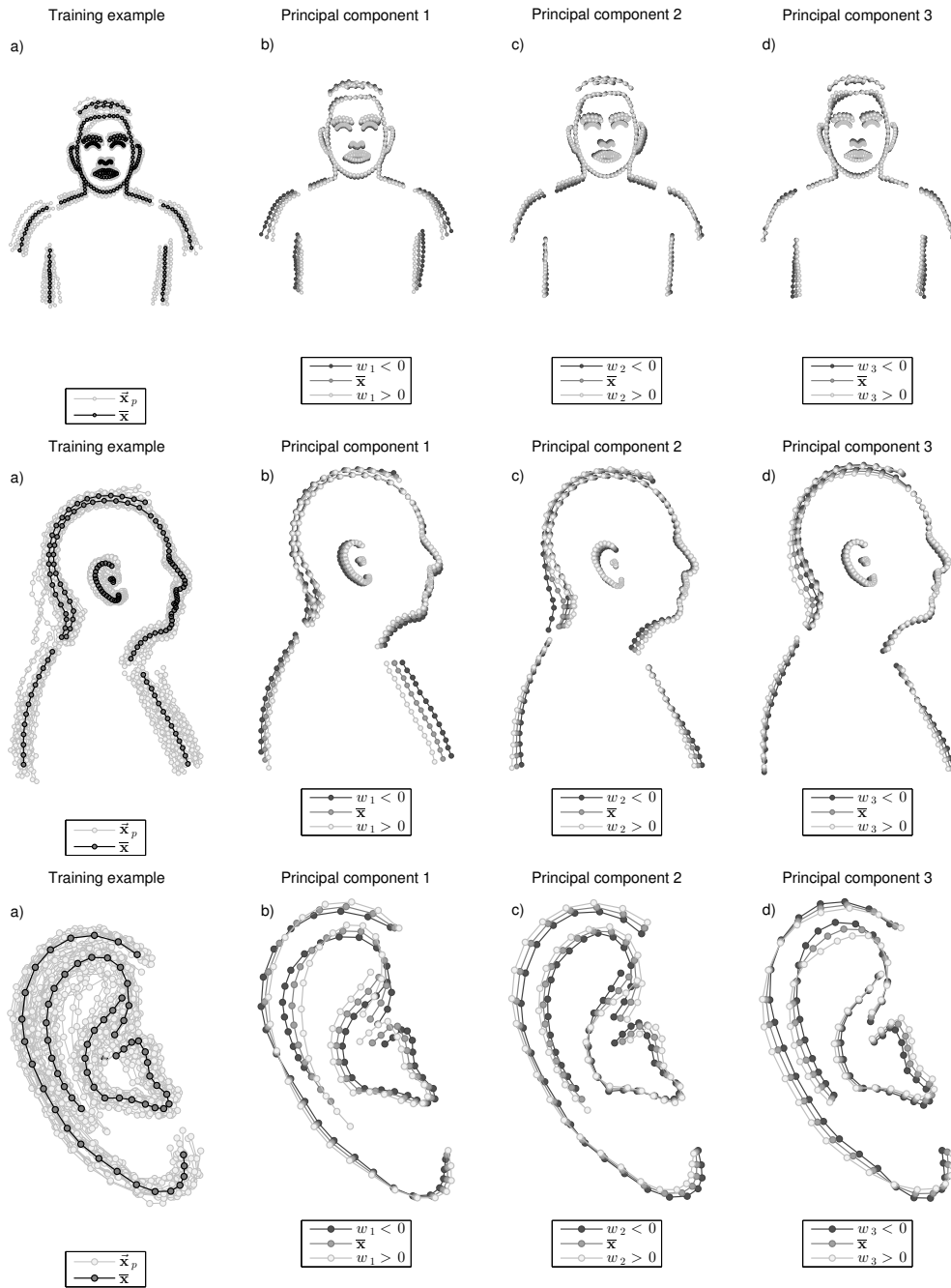


Figura 3.4: a) Vectores de entrenamiento, \vec{x}_p , y la forma promedio, \bar{x} , del modelo frontal (primera fila), modelo lateral (segunda fila), y modelo de la oreja (tercera fila). Las columnas b), c) y d) muestran la deformación de cada modelo, \bar{x} , al variar la ponderación, w_n , de cada una de tres componentes principales, en direcciones arbitrarias positivas ($w_n > 0$) y negativas ($w_n < 0$). La deformación acumulada de cada componente permite al modelo ajustarse a nuevos sujetos.

acuerdo a la información del entrenamiento. Para lograr esto, se recogen nuevos perfiles de gris, $\tilde{\mathbf{g}}_n^X$, de longitud M (con $M > L$), sobre las líneas perpendiculares al modelo \mathbf{X} en su posición actual. Estos perfiles se almacenan en la matriz $\mathbf{G}_{M \times N}^X$ y se comparan con los perfiles de gris del entrenamiento $\mathbf{G}_{L \times N}^p$ en busca de una posición en la cual ambos perfiles sean lo más similar posible. Para realizar la comparación se usa la métrica de la distancia de Mahalanobis [222], dada como:

$$D^2 = (\tilde{\mathbf{g}}_n^X - \bar{\mathbf{g}}^p)^T (\text{Cov}[G])^{-1} (\tilde{\mathbf{g}}_n^X - \bar{\mathbf{g}}^p) \quad (3.3)$$

donde $\bar{\mathbf{g}}^p$ es el promedio de $\mathbf{G}_{L \times N}^p$, $\tilde{\mathbf{g}}_n^X$ es el perfil de gris de largo L del punto n siendo comparado, tomado de $\mathbf{G}_{M \times N}^X$, y $\text{Cov}[G]$ es la matriz de covarianza de G , dada por

$$\text{Cov}[G] = \frac{1}{L-1} \mathbf{G}_{L \times N}^p{}^T \mathbf{G}_{L \times N}^p \quad (3.4)$$

Se comparan todas las posibles posiciones de los perfiles de entrenamiento dentro del perfil de gris actual. La recomendación es mover el punto a la posición para la cual D^2 sea mínima. De esta forma, el algoritmo entrega desplazamientos recomendados para todos los puntos del modelo, dentro del vector $d\mathbf{X} = (dx_1, dy_1, \dots, dx_n, dy_n)^T$. Así, la nueva posición sugerida para los puntos de \mathbf{X} es:

$$\mathbf{X}^t = \mathbf{X} + d\mathbf{X}. \quad (3.5)$$

Teniendo esta posición sugerida, el objetivo es ahora reubicar y deformar \mathbf{X} de forma tal que que minimice el error con respecto al objetivo \mathbf{X}^t . El primer paso es reubicar \mathbf{X} cambiando su posición \mathbf{x}_c , rotación θ y escala s . El error residual se reduce proyectando el error en las componentes principales, y calculando los coeficientes w_n que deformen \mathbf{X} para que coincida lo mejor posible con \mathbf{X}^t , minimizando el error.

El algoritmo continúa con la siguiente iteración, utilizando la forma final \mathbf{X}^f , escala s , rotación θ y posición \mathbf{x}_c final de la iteración actual como el estado inicial del modelo en la siguiente iteración. De esta forma, cada iteración entrega un modelo cada vez más cercano a la forma deseada. Las iteraciones continúan hasta que el modelo deja de deformarse en forma significativa, lo que significa que éste ha alcanzado un estado de convergencia. El algoritmo del ASM se puede estudiar con más detalle en las referencias [221, 223–226].

En la figura 3.5 se muestra un ejemplo del ajuste iterativo del modelo lateral tras 1, 9, 15 y 39 iteraciones. Los casos b), c) y d) muestran el modelo inicial \mathbf{X} en dichas iteraciones, también se muestran los perfiles de gris $\mathbf{G}_{M \times n}^X$ perpendiculares al modelo, el modelo objetivo \mathbf{X}^t que resulta de la comparación de perfiles de gris y sugiere al algoritmo una mejor ubicación de cada punto a lo largo del perfil de gris. Tras minimizar el error entre \mathbf{X} y \mathbf{X}^t , el algoritmo alcanza un estado final

para la iteración actual, \mathbf{X}^f , el cual se usa como estado inicial para la siguiente iteración.

El último caso (e), en la figura, muestra el modelo ajustado al sujeto de la imagen tras la convergencia (es decir, tras dejar de deformarse en forma significativa). Las diferencias que se observan entre el modelo ajustado y los bordes reales del sujeto en la fotografía serán analizadas y cuantificadas en la sección 3.2.

Este ejemplo de ajuste requirió un tiempo de procesamiento de 51 segundos en una computadora personal² y sobre una imagen de 15 mega píxeles. Este método resulta entonces muy conveniente para obtener la fotoantropometría en forma automática, como se explica en la siguiente sección.

3.1.5. Antropometría a partir del modelo ajustado

El modelo ajustado en la sección anterior permite la obtención de la antropometría del sujeto de una manera simple y automática. En este trabajo son de interés las medidas propuestas en la base de datos de CIPIC [20], las cuales se muestran en la figura 3.6.

Un programa en MATLAB obtiene la información antropométrica midiendo la distancia euclideana (en unidades de píxeles) entre puntos específicos del modelo ajustado; estos puntos, conocidos como puntos *somatométricos* [27] se muestran como pequeños círculos en la figura 3.6.

Se obtuvieron los siguientes parámetros antropométricos: x_1 ancho de la cabeza, x_2 altura de la cabeza, x_6 ancho del cuello, x_7 altura del cuello, x_9 ancho de la parte superior del torso, x_{10} altura a la parte superior del torso, y x_{12} ancho de los hombros, todos estos obtenidos a partir del modelo frontal. x_3 profundidad de la cabeza, x_4 desplazamiento vertical de la pinna a partir del centro de la cabeza, x_5 desplazamiento horizontal de la pinna a partir del centro de la cabeza, x_8 profundidad del cuello, y x_{11} profundidad de la parte superior del torso, sobre el modelo lateral. $d1$ altura de la *cavum conchae*³, $d2$ altura de la *cymba conchae*⁴, $d3$ ancho de la *cavum conchae*, $d4$ altura de la *fossa*⁵, $d5$ altura de la *pinna*⁶, $d6$ ancho de la *pinna*, y $d7$ ancho de la incisión *intertragal*, estos últimos sobre el modelo de la oreja.

Las distancias obtenidas, en unidades de píxeles, son convertidas a unidades

²Se empleó una PC de propósito general con un procesador Intel Core i5 1.7 GHz, 4 GB RAM, Mac OS X.

³Voz latina traducible como: cavidad de la concha.

⁴Voz latina, “*cymba*” se refiere a la cavidad angosta, con la forma de una balsa de remos.

⁵Voz latina para “fosa” o “depresión”.

⁶También conocida como “aurícula”, la *pinna* es una voz latina que puede traducirse al español como “ala” o “solapa”.

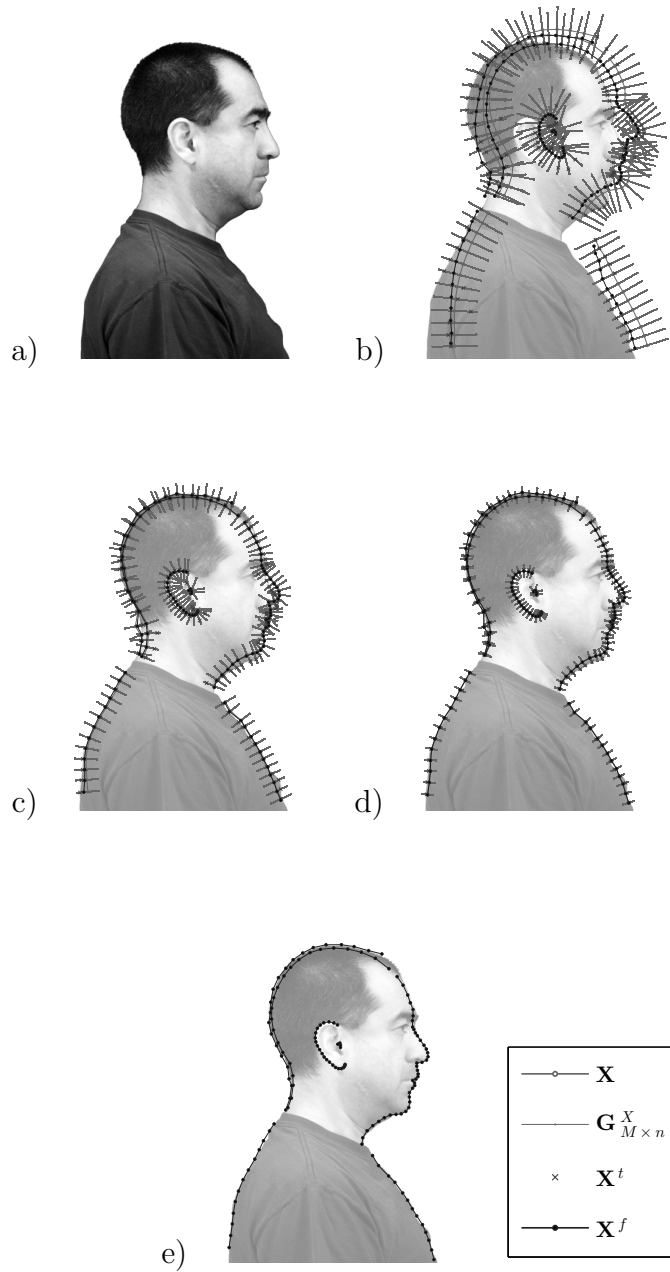


Figura 3.5: Ejemplo del proceso de ajuste del ASM. a) imagen original, b) primera iteración, c) novena iteración, d) quinceava iteración, y e) modelo final, ajustado tras 39 iteraciones.

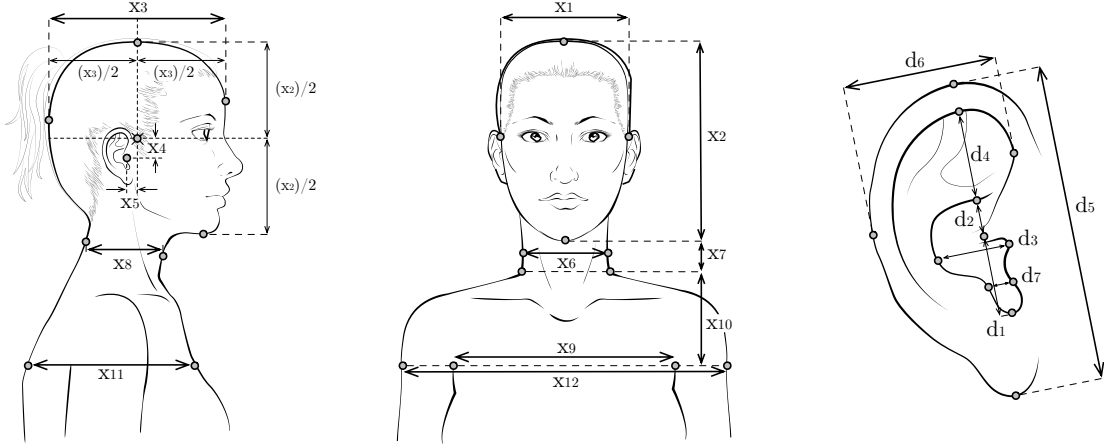


Figura 3.6: Parámetros antropométricos medidos sobre los modelos ajustados, propuestos originalmente en [20].

métricas usando el factor de conversión:

$$FC = \frac{1}{100} \cdot \frac{r_x \cdot s_2}{s_1 \cdot d_x}, \quad (3.6)$$

dado en unidades de px/cm (píxeles sobre centímetro), y donde r_x es la resolución o número de píxeles a lo largo del eje horizontal x del sensor de la cámara, d_x es la medida física (en metros) del sensor de la cámara a lo largo del mismo eje, s_1 y s_2 se obtienen a partir de la ecuación gaussiana de las lentes [227]:

$$\frac{1}{f} = \frac{1}{s_1} + \frac{1}{s_2} \quad (3.7)$$

y de la expresión:

$$D = s_1 + s_2, \quad (3.8)$$

donde f es la distancia focal de la lente y D es la distancia del sujeto al plano focal de la cámara.

Este factor de conversión es validado fotografiando una regla graduada, a la misma distancia en la que se retrata el sujeto (1.5 m). Para las lentes empleadas en este trabajo el factor de conversión fue de 78.27 px/cm para la lente de 50 mm, y 42.63 px/cm para la lente de 28 mm.

3.2. Validación

Se presentan a continuación las actividades realizadas para la validación del algoritmo ASM y del proceso de antropometría.

3.2.1. Validación del ASM

La validación del algoritmo de ajuste del ASM (presentado en la sección 3.1.4) se realizó usando una técnica de validación cruzada conocida como “*leave-one-out*” [228]. Esta técnica es particularmente útil en los casos en los que no es recomendable separar muchos datos para validación del grupo de entrenamiento, lo que tendría como consecuencia una disminución en la efectividad del entrenamiento.

Esta técnica de validación consiste en usar todos los datos para entrenamiento salvo uno, que se reserva para validar. El procedimiento se repite, entrenando con todos los datos, salvo uno distinto al del caso anterior, que será usado nuevamente para validación. Al final del procedimiento se habrán realizado tantas validaciones como el número de datos que se tenían originalmente.

La técnica se aplicó para el caso del ASM de la siguiente forma: se dispone de material fotográfico de once sujetos, así que se realizaron once sesiones de entrenamiento, cada una de ellas dejando fuera el material fotográfico de un sujeto distinto cada vez, y entrenando el ASM con los diez sujetos restantes.

Cargando cada sesión de entrenamiento, se realizó el ajuste del ASM sobre la fotografía del sujeto excluido, y se midió el error entre cada par de puntos respectivos entre el modelo ajustado por el algoritmo y el modelo anotado manualmente durante el entrenamiento. El error entre cada par de puntos es una distancia euclideana en unidades de pixel, estos se promediaron para todo el modelo y el resultado se convirtió a unidades de centímetros (usando el factor de conversión, ecuación 3.6).

En el cuadro 3.1 se muestra este error promedio del ajuste para cada uno de los once sujetos y para cada uno de los tres modelos. Los valores de error van desde 0.37 cm a 0.98 cm para el modelo frontal, de 0.37 cm a 1.04 cm para el modelo lateral, y de 0.13 cm a 0.38 cm para el modelo de la oreja. El error promedio para todos los sujetos es de 0.71 cm para el modelo frontal, 0.64 cm para el modelo lateral, y 0.23 cm para el modelo de la oreja.

Los efectos que estos errores pudieran introducir en la HRTF personalizada serán analizados en la sección 6.4.

3.2.2. Validación de la antropometría

La información antropométrica, obtenida utilizando el procedimiento descrito en la sección 3.1.5, fue validada comparando esos resultados con información antropométrica obtenida a partir de la anotación manual del *experto*⁷ sobre el mismo grupo de imágenes.

⁷Se describe como *experto* al operador humano que realiza una actividad, y que será reemplazado por un proceso automatizado en la computadora.

Sujeto	Error de cada modelo (cm)			Error promedio (cm)
	frontal	lateral	oreja	
1	0.72	0.48	0.30	0.50
2	0.61	0.92	0.38	0.63
3	0.98	0.71	0.20	0.63
4	0.67	0.68	0.33	0.56
5	0.75	0.58	0.15	0.49
6	0.37	0.81	0.13	0.44
7	0.70	0.50	0.16	0.45
8	0.71	0.37	0.26	0.45
9	0.60	0.57	0.21	0.46
10	0.91	1.04	0.16	0.70
11	0.83	0.41	0.21	0.48
Promedio	0.71	0.64	0.23	0.53
Desv. est.	0.16	0.21	0.08	0.09

Cuadro 3.1: Error promedio, en centímetros, de los modelos ajustados, relativos a los modelos anotados manualmente

Idealmente, los resultados de la foto-antropometría automatizada deberán ser iguales a los valores obtenidos a partir de las anotaciones del experto. Los resultados se muestran en el cuadro 3.2. Las dos primeras columnas indican el modelo y el parámetro antropométrico. La tercera columna muestra los valores de cada parámetro promediados entre los once sujetos, junto con sus desviaciones estándar (en paréntesis). La cuarta columna muestra el error absoluto promedio, en centímetros. Estos errores (y sus desviaciones estándar) son del orden de 0.1 cm para los parámetros de la oreja, menos de 0.6 cm para el modelo lateral, y menos de 1.33 cm para el modelo frontal.

Otra medida de error es el error relativo (porcentual) mostrado en la quinta columna; se obtiene como el cociente entre el error absoluto promedio y el valor promedio de cada parámetro antropométrico. Para el caso de la oreja, los valores de error van desde valores bajos como 1.12 % para $d5$, a valores altos como 23.75 % para $d7$. Esto puede explicarse al considerar que, aunque los valores de error absoluto promedio son de alrededor de 0.1 cm, los parámetros correspondientes abarcan un intervalo que se extiende de dimensiones grandes como 6.58 cm para $d5$, a dimensiones tan pequeñas como 0.58 cm para $d2$, entre ellas hay una diferencia de un orden de magnitud. Los resultados para el modelo lateral y frontal son similares: el error absoluto promedio más grande, 1.33 cm para x_{12} , solamente representa un 2.88 % de error relativo.

Dos parámetros, x_7 y x_5 , tienen errores relativos mayores a 100 %, por lo tanto serán excluidos del procedimiento de personalización de HRTF en secciones

posteriores de este trabajo. x_7 es la distancia entre la punta de la barbilla y la base del cuello. Esta distancia es muy variable como consecuencia de pequeños cambios en la postura de la cabeza del sujeto durante el proceso de fotografiado.

En el capítulo 9 se presentará un análisis de la influencia de estos errores en la personalización de las HRTF.

Modelo	Parámetro	Valor promedio cm (cm)	Error abs. promedio cm (cm)	Error relativo promedio
Frontal	x_1	14.52 (0.70)	0.63 (0.46)	4.41 %
	x_2	22.38 (1.41)	0.65 (0.58)	2.86 %
	x_6	10.64 (0.59)	0.56 (0.32)	5.27 %
	x_7	0.45 (0.23)	0.39 (0.30)	109.91 %
	x_9	29.53 (2.14)	0.70 (0.69)	2.34 %
	x_{10}	14.49 (0.94)	1.20 (0.81)	8.36 %
	x_{12}	43.53 (3.43)	1.33 (1.62)	2.88 %
Lateral	x_3	19.33 (0.87)	0.41 (0.32)	2.20 %
	x_4	2.38 (0.25)	0.27 (0.29)	15.08 %
	x_5	0.40 (0.55)	0.27 (0.17)	128.19 %
	x_8	11.22 (1.17)	0.28 (0.13)	2.59 %
	x_{11}	27.60 (2.37)	0.58 (0.47)	2.20 %
Oreja	d_1	1.65 (0.16)	0.11 (0.12)	6.59 %
	d_2	0.58 (0.12)	0.09 (0.07)	17.83 %
	d_3	1.64 (0.23)	0.13 (0.08)	7.74 %
	d_4	1.61 (0.49)	0.15 (0.13)	12.82 %
	d_5	6.58 (0.71)	0.07 (0.06)	1.12 %
	d_6	2.75 (0.26)	0.11 (0.09)	4.09 %
	d_7	0.68 (0.14)	0.16 (0.09)	23.75 %

Cuadro 3.2: Valores foto-antropométricos promedio, y sus errores. En paréntesis se muestran las desviaciones estándar.

Capítulo 4

Métodos de personalización de HRTF

A continuación se presentan tres métodos de personalización de HRTF a partir de la antropometría del sujeto. Los dos primeros se reproducen de la literatura pero con mejoras de desempeño, principalmente para admitir la fase de la HRTF. El tercer método se presenta como un método nuevo, basado en redes neuronales artificiales.

4.1. Mejor coincidencia antropométrica

Zotkin *et alii* [19] propusieron la personalización de HRTF usando la foto-antropometría del sujeto (al que a partir de ahora llamaremos *sujeto objetivo*) seleccionando las HRTF, dentro de la base de datos, del sujeto morfológicamente más parecido al sujeto objetivo. A este método lo llamaremos “mejor coincidencia antropométrica”; parte de la hipótesis de que dos sujetos con antropometría similar tendrán HRTF compatibles, que produzcan efectos de ubicación espacial de forma intercambiable. Esta hipótesis es, en principio, correcta, pues desde su definición, las HRTF son consecuencia de la antropometría.

Sin embargo, aún queda pendiente comprobar si las similitudes en términos de un grupo determinado de parámetros antropométricos y un grupo específico de sujetos en la base de datos aún producen HRTF compatibles, o hasta que grado la compatibilidad puede alcanzarse.

Para evaluar esta interrogante, se procedió de la siguiente forma: las HRTF complejas (magnitud y fase, en su representación con números complejos) de todos los sujetos de la base de datos se agruparon dentro de la matriz $\mathbf{H}_{S \times F}$, donde S es el número de sujetos, y F el número de componentes de frecuencia. Se tiene también

una matriz correspondiente con los datos antropométricos de los mismos sujetos, representada como $\mathbf{X}_{S \times P}$, donde P es el número de parámetros antropométricos tomados de la base de datos.

Las distribuciones de los parámetros antropométricos (columnas de X) se normalizan para tener promedio de cero y varianza unitaria, como se muestra en la figura 4.1. Es decir, a las distribuciones originales, ejemplificadas en la figura (4.1.a), se les extrae sus respectivos promedios μ_p , resultando en (4.1.b), y después se dividen entre su desviación estándar σ_p , resultando en las distribuciones normalizadas (4.1.c). Todo esto se hizo para evitar trabajar con datos ubicados en rangos numéricos tan distintos, que van desde unos cuantos milímetros para los parámetros de la oreja, hasta medio metro para el caso del ancho de los hombros.

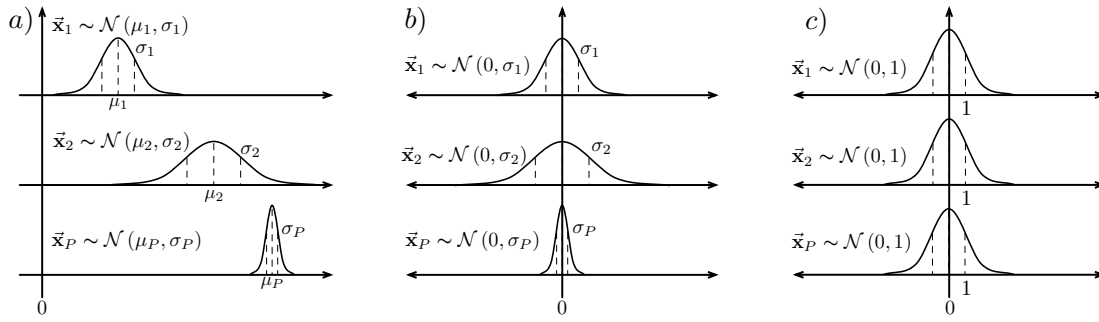


Figura 4.1: Normalización de las distribuciones de los parámetros antropométricos a) ejemplo de las distribuciones originales, b) tras sustraer los promedios respectivos μ_p , y c) tras dividir entre sus desviaciones estándar σ_p .

Tras normalizar las distribuciones, se selecciona un sujeto cualquiera de la base de datos de CIPIC, denominado *sujeto objetivo* t , y sus datos se remueven tanto de \mathbf{X} como de \mathbf{H} .

La información antropométrica $\bar{\mathbf{x}}_p^t$ del sujeto t se compara con la antropometría de todos los sujetos restantes de la base de datos (las filas de \mathbf{X}), en busca de la mejor coincidencia antropométrica. La métrica usada para la comparación es el error (distancia) cuadrático medio ponderado entre elementos de los vectores normalizados de antropometría, y se obtiene con la siguiente expresión:

$$d_s = \sqrt{\frac{1}{N_\alpha} \sum_{p=1}^P \alpha_p^2 |x_p^t - x_p^s|^2}, \quad (4.1)$$

donde x_p^t es el p -ésimo elemento del vector de antropometría del sujeto objetivo, x_p^s es el p -ésimo elemento de la s -ésimo vector fila de la matriz $\mathbf{X}_{(S-1) \times P}$; $N_\alpha = \sum_p \alpha_p^2$ es un factor de normalización promedio ponderado, que es igual a $N_\alpha = P$ (número de parámetros antropométricos) cuando $\alpha_p = 1$, y α_p son las

ponderaciones de importancia relativa de cada parámetro antropométrico (estos valores fueron calculados con el procedimiento que se describe en el capítulo 8), y son, para el oído derecho y la dirección frontal ($\theta = \phi = 0^\circ$):

$$\begin{aligned} x_1 &= 18\%, & x_6 &= 17.8\%, & x_{11} &= 34.6\%, & d_3 &= 38.9\%, & d_7 &= 5.7\%. \\ x_2 &= 32.7\%, & x_8 &= 3.7\%, & x_{12} &= 6.2\%, & d_4 &= 20.7\%, \\ x_3 &= 18.2\%, & x_9 &= 22.2\%, & d_1 &= 15\%, & d_5 &= 22\%, \\ x_4 &= 14.5\%, & x_{10} &= 7.7\%, & d_2 &= 43.2\%, & d_6 &= 14.7\%, \end{aligned}$$

Así, la mejor coincidencia antropométrica es el sujeto s dentro de la base de datos para el cual la ecuación (4.1) entrega el valor mínimo.

En este punto del procedimiento ya se conoce qué sujeto dentro de la base de datos es antropométricamente más parecido al sujeto objetivo; el siguiente paso es verificar si existe relación entre las distancias antropométricas y los errores entre sus antropometrías. Esto se presenta a continuación.

La HRTF del sujeto s de la base de datos, $\vec{\mathbf{h}}^s$ (fila s de la matriz \mathbf{H}) se compara con la HRTF del sujeto objetivo $\vec{\mathbf{h}}^t$, y la diferencia entre ellas se cuantifica con la métrica de *distorsión espectral* (SD¹), dada por la siguiente ecuación, modificada de la referencia [22]:

$$\text{SD} = \sqrt{\frac{1}{N_A} \sum_{f=1}^F A_f^2 \left[20 \log_{10} \frac{|h_f^s|}{|h_f^t|} \right]^2}, \quad (4.2)$$

donde SD está dada en decibeles, F es el número de componentes de frecuencia de la HRTF, y h_f^s y h_f^t son el f -ésimo elemento de los vectores $\vec{\mathbf{h}}^s$ y $\vec{\mathbf{h}}^t$, respectivamente. A_f son los factores de ponderación de la importancia porcentual relativa de cada componente de frecuencia de la HRTF, y para este caso se toman como:

$$A_f = 10^{\Delta L_f/10}, \quad (4.3)$$

de acuerdo a la “curva de ponderación A” usada habitualmente en mediciones de nivel de sonido [229, 230], y $N_A = \sum_f A_f^2$, que resulta en $N_A = F$ cuando $A_f = 1$.

Se espera que la distorsión espectral de la HRTF que corresponde a la mejor coincidencia antropométrica, también sea un mínimo (o cerca del mínimo) si se compara con las distorsiones espectrales obtenidas a partir de los demás sujetos de la base de datos. La certeza de esta suposición será evaluada exitosamente en la sección 6.1.

¹SD, siglas del inglés *spectral distortion*.

4.2. Regresión lineal múltiple

El siguiente método de personalización de HRTF es la síntesis a través de una combinación lineal de una base ortogonal obtenida a partir de la biblioteca mediante el análisis de componentes principales [24], y dictada por el resultado de una regresión lineal múltiple entre la antropometría y la base ortogonal. Este método, que denominaremos *regresión lineal múltiple* (MLR²), ha sido empleado, con algunas variaciones, por distintos autores como Nishino *et al.* [22, 202, 203], Hu *et al.* [205], Xu *et al.* [164, 208] y Hugeng *et al.* [209, 210, 217, 231].

En este trabajo se presentan dos tipos de regresión, parten de la misma idea, pero con una diferencia fundamental ubicada en la inclusión o exclusión de la fase; los métodos son:

- (a) Regresión de la magnitud logarítmica del espectro de la HRTF (tal y como se ha presentado en la literatura).
- (b) Regresión del espectro logarítmico complejo de la HRTF, que incluye la fase de la HRTF y la información del tiempo interaural.

Se probaron otros tipos de regresión (en escala lineal y logarítmica, en representación “magnitud, fase”, y “parte real, parte imaginaria”, etc.), pero ninguno de ellos resultó ser más efectivo y no se reportan a continuación.

Método

Antes de la realizar la regresión lineal, los vectores $\tilde{\mathbf{h}}_F^s$ de HRTF de la base de datos son tratados de la siguiente manera. Para el caso (a), la fase de descarta, solamente se conserva la magnitud del espectro, es decir:

$$\tilde{\mathbf{h}}_F^s \leftarrow \left| \tilde{\mathbf{h}}_F^s \right|. \quad (4.4)$$

Para el caso (b) se aplica una corrección a la fase de cada elemento de frecuencia de todas las HRTF (de los S sujetos), esto se hace con el propósito de evitar, en lo posible, discontinuidades en las distribuciones de fase al pasar éstas de $-\pi$ a $+\pi$. Las correcciones de fase se almacenan para restituir las fases originales al final del procedimiento.

En ambos casos (a) y (b), los vectores de HRTF se convierten a escala logarítmica con la expresión:

$$\tilde{\mathbf{h}}_F^s \leftarrow 20 \cdot \log_{10}(\tilde{\mathbf{h}}_F^s). \quad (4.5)$$

²MLR, siglas del inglés *Multiple Linear Regression*.

Para el caso (b) esto resulta en un logaritmo complejo.³

Después se obtiene la HRTF promedio de las HRTF de todos los sujetos:

$$\bar{\mathbf{h}}_F = \frac{1}{S} \sum_{s=1}^S \tilde{\mathbf{h}}_F^s, \quad (4.6)$$

y a todas las HRTF se les resta esta HRTF promedio:

$$\hat{\mathbf{h}}_F^s \leftarrow \tilde{\mathbf{h}}_F^s - \bar{\mathbf{h}}_F. \quad (4.7)$$

Los vectores que resultan, $\hat{\mathbf{h}}_F^s$, se agrupan dentro de la matriz $\hat{\mathbf{H}}_{S \times F}$. Y después se realiza el análisis de componentes principales (PCA), es decir, se obtienen los *vectores propios y valores propios*⁴ de la matriz de covarianza

$$\mathbf{C}_{F \times F} = \frac{1}{S-1} \hat{\mathbf{H}}_{S \times F}^T \hat{\mathbf{H}}_{S \times F}. \quad (4.8)$$

Los vectores propios se ordenan dentro de las columnas de la matriz $\mathbf{E}_{F \times F}$, en forma descendente de acuerdo con sus valores propios respectivos.

En este punto las HRTF pueden sintetizarse como una combinación lineal de componentes (principales), cada una con un coeficiente (de ponderación) lineal específico. Estos coeficientes se determinan al relacionar la información antropométrica con las ponderaciones de las componentes principales, a través de una regresión lineal múltiple (MLR) dada por:

$$\mathbf{A}_{P+1 \times N} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \quad (4.9)$$

donde $\mathbf{A}_{P+1 \times N}$ es la matriz de coeficientes de regresión. \mathbf{X} es la matriz que contiene los datos antropométricos de los sujetos de la base de datos. En este método \mathbf{X} tiene dimensiones $S \times (P+1)$, pues la primera columna debe llenarse con “1”, lo que da lugar a un coeficiente constante independiente en la regresión lineal. \mathbf{W} es la matriz de ponderaciones que relaciona las componentes principales con las HRTF, se obtiene de:

$$\mathbf{W}_{S \times N} = \mathbf{E}^{-1} \hat{\mathbf{H}}^T. \quad (4.10)$$

Para realizar la síntesis de la HRTF del sujeto objetivo t se requiere de sus datos antropométricos, agrupados dentro del vector $\vec{\mathbf{x}}_{P+1}^t$ (el primer elemento también debe ser “1”, es decir $x_1^t = 1$). Este vector se multiplica por la matriz de coeficientes de regresión para obtener los coeficientes lineales:

$$\vec{\mathbf{w}}_N = \vec{\mathbf{x}}_{P+1}^{tT} \mathbf{A}, \quad (4.11)$$

³El logaritmo de un número complejo z es: $\log z = \log |z| \cdot e^{i\theta}$, donde $|z|$ es la magnitud y θ el ángulo de z .

⁴Referidos como *eigenvectors* y *eigenvalues* en la literatura en inglés.

y después se combinan las componentes principales con los coeficientes lineales, en forma lineal, para obtener:

$$\hat{\mathbf{h}}_F^{tT} = \mathbf{E}\bar{\mathbf{w}}^T. \quad (4.12)$$

En este punto se agrega la HRTF promedio (que se había restado a todos los vectores al inicio del procedimiento),

$$\mathbf{h}_F^t = \bar{\mathbf{h}}_F + \hat{\mathbf{h}}_F^t. \quad (4.13)$$

La HRTF resultante, $\vec{\mathbf{h}}_F^t$, se convierte a escala lineal con la expresión:

$$\vec{\mathbf{h}}_F^t \leftarrow 10^{(\vec{\mathbf{h}}_F^t/20)}, \quad (4.14)$$

y, para el caso (b), se restituye la fase original a cada uno de las componentes de frecuencia.

Para los dos casos, (a) y (b), el vector $\vec{\mathbf{h}}_F^t$ es la HRTF sintetizada para el sujeto objetivo t cuya antropometría está contenida en el vector $\vec{\mathbf{x}}^t$.

4.3. Síntesis con red neuronal artificial

El problema de la personalización de HRTF a partir de la antropometría también se abordó usando redes neuronales artificiales (ANN⁵) [232–234], que son apreciadas en la comunidad académica por su capacidad de ajustar funciones no lineales.

Las redes neuronales artificiales han sido usadas anteriormente para propósitos de interpolación de HRTF [147, 154, 235]; en este trabajo se propone su uso para personalización de HRTF, utilizándolas como ajustadores no lineales, a diferencia del trabajo de Hu [213], quien propone su uso como ajustadores lineales.

La red neuronal recibe como información en su entrada al vector de información antropométrica $\vec{\mathbf{x}}_P$, y produce un vector de HRTF $\vec{\mathbf{h}}_F$ en su salida. Para su entrenamiento se utiliza la información de la base de datos, tanto la matriz de antropometría \mathbf{X} , como la matriz de HRTF correspondiente, \mathbf{H} .

El primer paso de determinar la arquitectura de la red más adecuada a las dimensiones del problema, y que produzca el mínimo error alcanzable, cuantificado en términos de la distorsión espectral. Se probaron dos arquitecturas de redes, ambas sin retroalimentación: una red neuronal de *alimentación frontal* (FFNN⁶), y una red neuronal de *cascada frontal* (CFNN⁷), esta última caracterizada por tener conexiones de cada cada capa con todas las capas subsecuentes.

⁵ANN, siglas del inglés *artificial neural networks* (ANN).

⁶FFNN, siglas del inglés *feed forward neural network*.

⁷CFNN, siglas del inglés *cascade forward neural network*.

Para cada arquitectura se probaron varias configuraciones en términos del número de capas ocultas, funciones de transferencia y número de neuronas en cada capa. Las funciones de transferencia que se probaron fueron: la función tangente sigmoidea hiperbólica (no lineal), la función lineal, usada generalmente para la capa de salida, la función de base radial, la función escalón, y la función lineal saturada [233, 234].

La evaluación de cada configuración y arquitectura se realizó entrenando una red con la base de datos de CIPIC. Para el entrenamiento se usó el algoritmo de Levenberg–Marquardt, con una repartición de datos de 70% para el entrenamiento, 15% para validación, y 15% para pruebas. Después, se utilizó la técnica *leave-one-out*, dejando fuera un sujeto objetivo, y se utilizó el resto de la información para entrenar la red. La red entrenada se alimenta entonces con la información antropométrica del sujeto objetivo, y la HRTF que produce la red es comparada con la HRTF original del sujeto t (inicialmente contenida en la base de datos), esta comparación se hace en términos de la distorsión espectral. El procedimiento se repite, dejando fuera en cada ocasión a cada uno de los 35 sujetos, y al final se obtiene el valor de distorsión espectral promedio de todos ellos.

Tras pruebas extensivas, se determinó que el mejor resultado se obtiene con una red neuronal de alimentación frontal (FFNN) con una capa oculta de 50 neuronas, y la función de transferencia tangente sigmoidea para la capa oculta y la función lineal para la capa de salida.

La red neuronal empleada se muestra en la figura 4.2, en ésta x_1, x_2, \dots, x_n , son los elementos del vector de datos antropométricos, \vec{x}_P , y h_1, h_2, \dots, h_F son los elementos de la HRTF resultante, \vec{h}_F . P es el número de parámetros antropométricos (entradas), y F es el número de componentes de frecuencia del vector de HRTF (salidas). Las matrices de coeficientes $\mathbf{W}_{50 \times P}^1$, $\mathbf{W}_{F \times 50}^2$ y los *vectores de sesgos*⁸ $\tilde{\mathbf{b}}_{50}^1$, $\tilde{\mathbf{b}}_F^2$ aplicados a las entradas y a las neuronas, se determinan durante la etapa de entrenamiento.

Este método de personalización de HRTF se empleó para predecir dos casos de HRTF:

- (a) magnitud logarítmica del espectro de la HRTF, y
- (b) espectro complejo logarítmico de la HRTF.

Previo al entrenamiento de la red, las HRTF de la base de datos se trataron de la misma forma que para el método anterior (ver sección 4.2), y para el caso (b) se entrenaron dos redes neuronales en paralelo, una para predecir la parte real del espectro, $\text{Re}(20 \log_{10} \mathbf{H})$, y la otra para la parte imaginaria, $\text{Im}(20 \log_{10} \mathbf{H})$.

Tras el entrenamiento, la red neuronal es usada para personalizar la HRTF de cualquier sujeto, dada su antropometría.

⁸En la literatura en inglés se usa el término *bias vectors*.

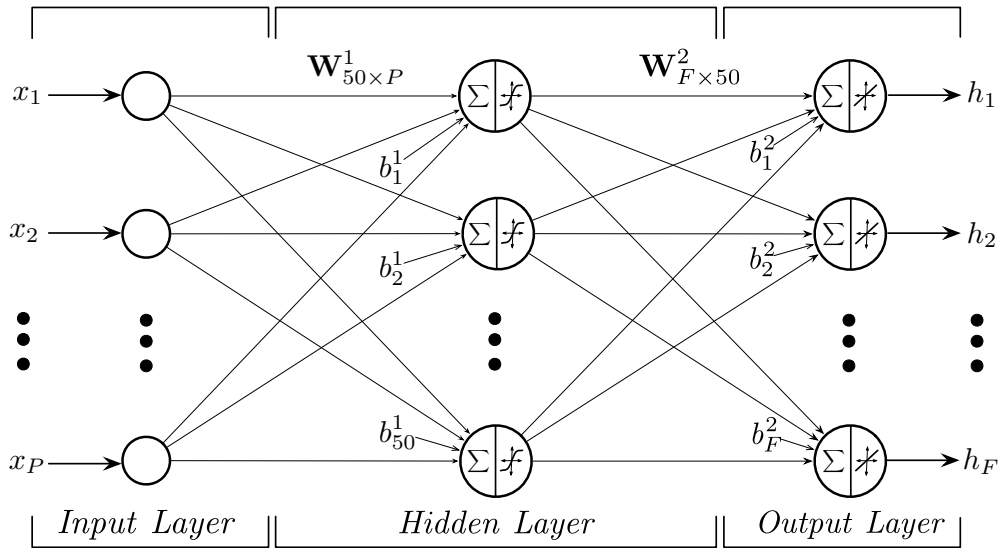


Figura 4.2: Red neuronal empleada, de alimentación frontal y con una capa oculta, que relaciona la información antropométrica con las HRTF.

Para el caso (a) la HRTF resultante $\vec{\mathbf{h}}_F$ se convierte a escala lineal. Para el caso (b) la HRTF personalizada se construye mediante la combinación de la parte real (producto de la primera red) y la parte imaginaria (de la segunda red), es decir:

$$\vec{\mathbf{h}}_F = \vec{\mathbf{h}}_F^{\text{Real}} + i \cdot \vec{\mathbf{h}}_F^{\text{Imaginaria}}; \quad (4.15)$$

ésta se convierte a escala lineal, y se le restituye la fase original.

Capítulo 5

Distorsión de espectro complejo

La métrica de la distorsión espectral (ecuación 4.2)

$$SD = \sqrt{\frac{1}{N_A} \sum_{f=1}^F A_f^2 \left[20 \log_{10} \frac{|h_f^s|}{|h_f^t|} \right]^2}, \quad (4.2)$$

cuantifica la diferencia de la raíz cuadrática media entre dos magnitudes de espectro, en decibeles. Esta métrica fue introducida para evaluar los resultados de personalización de HRTF por Nishino *et alii* [22] (ecuación 16), pero sin incluir la ponderación en frecuencia, es decir:

$$SD = \sqrt{\frac{1}{F} \sum_{f=1}^F \left[20 \log_{10} \frac{|h_f^s|}{|h_f^t|} \right]^2}, \quad (5.1)$$

sin embargo esta métrica tiene una deficiencia importante: ignora las diferencias de fase de los espectros comparados.

En los métodos presentados en el capítulo anterior se trabajó con dos casos; los errores del caso (a), magnitud del espectro, pueden ser cuantificados por la distorsión espectral (ecuación 4.2), sin embargo esto no es posible para el caso (b), en el que se trabajó con un espectro complejo.

Por lo tanto, se requiere una nueva métrica que tome en cuenta tanto las diferencias de magnitud, como las diferencias de fase. Se propone una solución para este problema a continuación:

Para comenzar, el argumento del logaritmo de la ecuación 4.2 se reordena algebraicamente, lo que resulta en:

$$SD = \sqrt{\frac{1}{N_A} \sum_{f=1}^F A_f^2 \left[20 \log_{10} \left(1 + \frac{|h_f^s| - |h_f^t|}{|h_f^t|} \right) \right]^2}. \quad (5.2)$$

En forma similar a esta ecuación, pero tomando en cuenta la magnitud y la fase del espectro, proponemos una nueva métrica, denominada *distorsión de espectro complejo* (CSD¹):

$$\text{CSD} = \sqrt{\frac{1}{N_A} \sum_{f=1}^F A_f^2 \left[20 \log_{10} \left(1 + \frac{|h_f^s - h_f^t|}{|h_f^t|} \right) \right]^2}, \quad (5.3)$$

que produce un valor de distorsión dado en decibeles. Para este caso, h_f^s y h_f^t son la f -ésima componente de frecuencia de las funciones de transferencia complejas $\vec{\mathbf{h}}^s$ y $\vec{\mathbf{h}}^t$. La diferencia compleja entre ambas (que involucra magnitud y fase), se calcula en el argumento del logaritmo como:

$$|h_f^s - h_f^t|. \quad (5.4)$$

Esta nueva métrica se puede entender mejor si el argumento del logaritmo se reordena algebraicamente a la forma:

$$\frac{|h_f^t| + |h_f^s - h_f^t|}{|h_f^t|}; \quad (5.5)$$

como se ilustra en la figura 5.1, y comparamos este argumento con el de la distorsión espectral, que es:

$$\frac{|h_f^s|}{|h_f^t|}. \quad (5.6)$$

Entonces, la SD es una medida logarítmica (en valor absoluto) de la relación entre los segmentos $|h_f^s|$ y $|h_f^t|$ (que puede ser mayor o menor a la unidad); mientras que la CSD es una medida logarítmica (siempre positiva) de la relación entre $|h_f^t| + |h_f^s - h_f^t|$ y $|h_f^t|$, que siempre resulta mayor que la unidad.

Es necesario mencionar que la SD y la CSD no siempre están relacionadas entre sí, se pueden encontrar casos en los que $\text{SD} < \text{CSD}$ y otros en los que $\text{SD} > \text{CSD}$. Este comportamiento (un tanto anti intuitivo) es consecuencia del cuadrado del logaritmo, que tiene una presencia peculiar desde la definición de la SD. Esto puede permitir que el logaritmo negativo de un cociente $|h_f^s|/|h_f^t|$ pequeño, menor a la unidad, resulte en $\text{SD} > \text{CSD}$, pese al hecho de que geoméricamente siempre se cumpla que:

$$|h_f^t| + |h_f^s - h_f^t| \geq |h_f^t|, \quad (5.7)$$

lo que falsamente sugiere que $\text{CSD} \geq \text{SD}$; este no siempre es el caso.

¹CSD, siglas del inglés *complex spectrum distortion*.

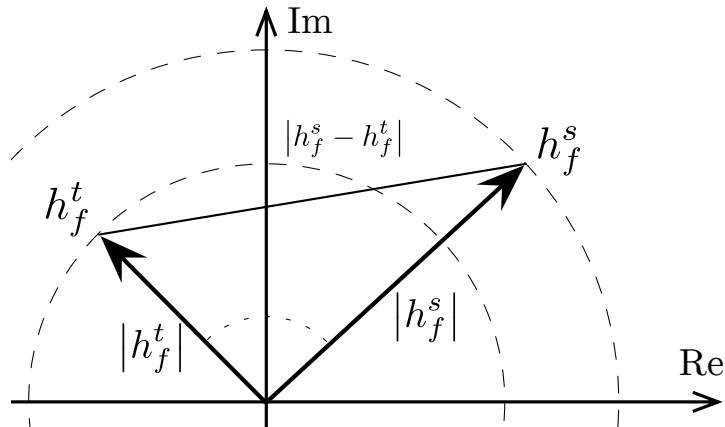


Figura 5.1: Ejemplo de las diferencias geométricas entre las componentes de h_f^t y h_f^s en el plano complejo.

Es importante resaltar que tanto la SD, como la CSD, son métricas monoaurales; al caracterizar el error en las HRTF de cada uno de los dos oídos del sujeto, es necesario utilizarlas por separado. Tampoco cuantifican otros aspectos de las HRTF como la diferencia de tiempo interaural (ITD), diferencia de nivel interaural (ILD), etc.

Capítulo 6

Validación de los métodos

Una vez que se presentaron las métricas para cuantificar las diferencias entre las HRTF originales y las personalizadas (SD y CSD, ver capítulo 5), el siguiente paso es la validación de los métodos de personalización de HRTF (presentados en el capítulo 4); este procedimiento se presenta y discute en este capítulo.

6.1. Mejor coincidencia antropométrica

Como se discutió en la sección 4.1, en lo que respecta al método de personalización mediante la búsqueda de la mejor coincidencia antropométrica dentro de la base de datos, se espera que la distorsión espectral, CSD, de la HRTF que corresponde a la mejor coincidencia antropométrica también sea un mínimo (o cerca del mínimo) si se compara con las distorsiones espectrales obtenidas con respecto a los demás sujetos de la base de datos.

Para evaluar este punto, se calculó la SD entre el sujeto objetivo y todos los sujetos de la base de datos, y los resultados se ordenaron de acuerdo a sus distancias antropométricas (de menor a mayor). Después, se empleó la técnica de *leave-one-out*, repitiendo el procedimiento anterior hasta dejar fuera a cada uno de los 35 sujetos de la base de datos como sujeto objetivo.

El promedio de todos los resultados se muestra en la figura 6.1. El análisis se hizo para cuatro bandas de frecuencia distintas, desde 0 Hz y con límites superiores de frecuencia de 3.4 kHz, 8 kHz, 17 kHz y 22.05 kHz. La primera banda corresponde al ancho de banda de comunicaciones telefónicas, de particular interés para investigación de telecomunicaciones binaurales. La segunda y la cuarta banda de frecuencia son las que reporta Nishino en su trabajo¹ [22]. La tercera banda

¹ La base de datos de HRTF que usa Nishino, sin embargo, no es la de CIPIC, sino una

representa razonablemente bien el intervalo audible de una persona de 10 años con audición normal [236]. En la figura se muestran las tres primeras bandas, la cuarta banda se omite de la gráfica por claridad, pues es muy similar a la tercera banda y se traslapa.

Los resultados se muestran ordenados de menor a mayor distancia antropométrica, es decir, desde los sujetos más parecidos (izquierda) a los menos parecidos (derecha).

Los valores de distorsión espectral fueron ajustados por una recta (también mostrada en la figura), y ésta resultó con pendiente positiva. Esto apoya la hipótesis de que la distorsión aumenta proporcionalmente con la distancia antropométrica, es decir, entre más parecidos sean los sujetos, menor es el error de distorsión espectral, y entre mayores sean las diferencias antropométricas, mayor es el error de distorsión espectral.

Pese a la proporcionalidad encontrada tras el ajuste lineal de los datos, la distorsión espectral se comporta en general en una forma un tanto errática. El coeficiente de correlación de Pearson es relativamente bajo, $\rho_{xy}^2 = 0.57, 0.52, 0.57$, para las tres bandas mostradas en la figura, respectivamente. Este resultado sugiere que los parámetros antropométricos o el número de sujetos en la base de datos son quizás insuficientes o inadecuados para propósitos de personalización de HRTF utilizando este método de mejor coincidencia antropométrica. Respecto a este punto, análisis adicional será presentado más adelante.

6.2. Comparación de los tres métodos

Los tres métodos de personalización de HRTF descritos en el capítulo 4 fueron evaluados y comparados cuantitativamente usando la técnica de *leave-one-out* y la base de datos de CIPIC. La evaluación se llevó a cabo para el caso del oído derecho y la dirección frontal (ángulo $\theta = \phi = 0^\circ$) y los parámetros antropométricos mostrados en la figura 3.6 (página 23), pero dejando fuera tanto x_5 como x_7 (como se explicó en la sección 3.2.2).

Un sujeto objetivo t se excluyó de la base de datos y se realizó la personalización de la HRTF para el sujeto t a partir de los datos de los sujetos restantes. Se calculó la distorsión espectral (SD, ecuación 5.2) y la distorsión de espectro complejo (CSD, ecuación 5.3) entre la HRTF personalizada para el sujeto objetivo t y la HRTF original del mismo sujeto, originalmente contenida en la base de datos. Este procedimiento se repitió dejando fuera cada uno de los 35 sujetos de la base de datos. Todos los valores de distorsión resultantes fueron promediados, obteniéndose así un valor de distorsión promedio para cada uno de los tres

propietaria, muestreada a 48 kHz, así que en realidad reportan valores para la banda de 0 a 24 kHz, y usan parámetros antropométricos diferentes a los de la base de datos de CIPIC

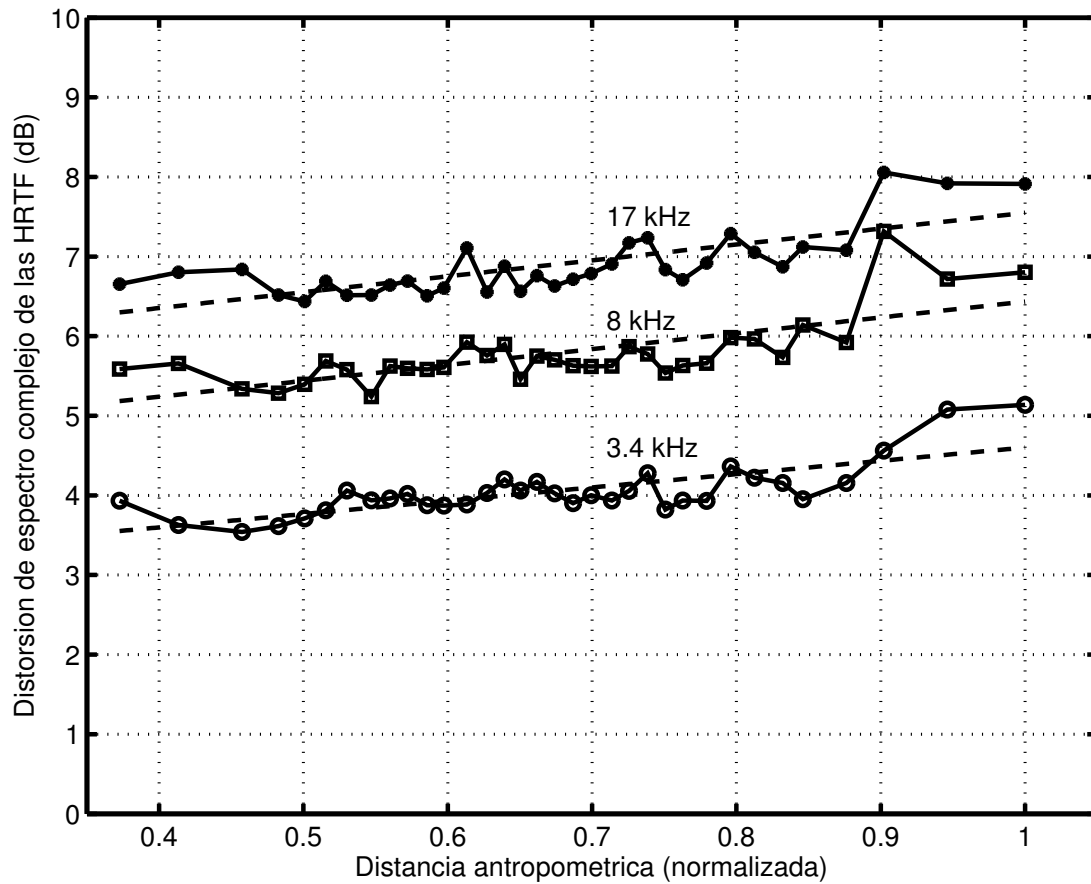


Figura 6.1: Distancias antropométricas promedio comparadas con sus respectivas distorsiones espectrales complejas promedio, medidas entre el sujeto objetivo y cada uno de los sujetos restantes de la base de datos de CIPIC. Se muestran los resultados para tres límites superiores de frecuencia: 3.4, 8 y 17 kHz.

métodos y cada una de las métricas (SD y CSD).

El análisis se hizo para las cuatro bandas de frecuencia descritas en la sección anterior, todas comenzando desde 0 Hz, y con límites superiores de frecuencia de 3.4 kHz, 8 kHz, 17 kHz, y 22.05 kHz.

Los valores promedio de distorsión entre la HRTF personalizada y la original, obtenidos con cada método y para cada ancho de banda se muestran en el cuadro 6.1 para el caso de la distorsión espectral (SD), y en el cuadro 6.2 para el caso de la distorsión de espectro complejo (CSD).

Distorsión espectral SD	Límite sup. (kHz)			
	3.4	8	17	22
1. Mejor coincidencia antropométrica:	3.3	4.6	6.0	6.2
2. Regresión lineal múltiple de la				
a) magnitud del espectro:	3.5	4.9	6.7	6.9
3. Síntesis con red neuronal de la				
a) magnitud del espectro:	6.1	11.2	12.8	14.8
4. Promedio entre pares de CIPIC:	3.4	5.1	6.5	6.7
5. HRTF del maniquí KEMAR:	2.9	5.0	6.7	7.0
6. Distancia a la HRTF promedio				
a) magnitud de espectro, promedio:	2.4	3.6	4.6	4.8

Cuadro 6.1: Distorsión espectral en decibelios, promediados sobre la base de datos de CIPIC entre la HRTF personalizada y la original, para los tres métodos de inferencia (números 1 a 3). Otros valores de distorsión espectral promediados sobre la base de datos de CIPIC (números 4 a 6), como referencia de comparación.

Distorsión espectral

Primero analizamos los resultados de la distorsión espectral (cuadro 6.1). Entre los tres métodos (números 1 a 3 en el cuadro), el método que se desempeña mejor, es decir, el que tiene valores menores de SD, es el método de *mejor coincidencia antropométrica* (1). Le sigue el método de *regresión lineal múltiple* con valores de distorsión mayores.

El método de síntesis con *red neuronal artificial* se desempeña peor que los otros métodos. Esto es desalentador, pues se tenía la expectativa de que la capacidad de la red neuronal de ajustar funciones no lineales permitiría describir mejor las relaciones entre la antropometría y las HRTF en forma mucho mejor que el caso de la regresión lineal.

Distorsión de espectro complejo CSD	Límite sup. (kHz)			
	3.4	8	17	22
1. Mejor coincidencia antropométrica:	3.9	5.6	6.7	6.9
2. Regresión lineal múltiple del b) espectro complejo:	4.1	5.9	7.2	7.3
3. Síntesis con red neuronal del b) espectro complejo:	6.5	8.9	10.2	10.4
4. Promedio entre pares de CIPIC:	4.1	5.9	7.0	7.2
5. HRTF del maniquí KEMAR:	3.1	4.3	5.5	5.6
6. Distancia a la HRTF promedio b) espectro complejo, promedio:	3.0	5.7	6.8	11.2

Cuadro 6.2: Distorsión de espectro complejo en decibeles, promediados sobre la base de datos de CIPIC entre la HRTF personalizada y la original, para los tres métodos de inferencia (números 1 a 3). Otros valores de distorsión de espectro complejo promediados sobre la base de datos de CIPIC (números 4 a 6), como referencia de comparación.

Distorsión de espectro complejo

En lo que respecta a la métrica de la distorsión de espectro complejo (CSD), cuadro 6.2, el método de mejor coincidencia antropométrica nuevamente se desempeña con los menores niveles de error de entre los tres métodos. Le sigue el método de la regresión lineal múltiple. Nuevamente, el método de red neuronal artificial es el de peor desempeño.

6.3. HRTF promedio y genérica

Las evaluaciones realizadas con las dos métricas de distorsión (SD y CSD) sugieren que el mejor método para personalizar HRTF (de entre los tres métodos comparados) es el de *mejor coincidencia antropométrica* (sección 4.1), esto se justifica al observar que este método ofrece los valores de distorsión más bajos para todos los casos.

Sin embargo, estos valores de distorsión aún deben demostrar ser lo suficientemente bajos, de forma tal que provean de una sensación espacial correcta en aplicaciones de tecnología binaural. Al respecto, Nishino [22, 148] postula un límite superior de distorsión espectral de 5.7 dB, y afirma que este nivel es suficiente para efectos “burdos” de ubicación espacial. Nishino trabaja con una base de datos propietaria de 72 sujetos y reporta valores de distorsión espectral por debajo de este umbral, es decir 4.0 dB para la banda de 0-8 kHz, y de esta forma valida

su método. Los resultados presentados en este trabajo son ligeramente mayores, pero del mismo orden de magnitud que los presentados por Nishino, pese a haber sido obtenidos con una base de datos distinta y con un menor número de sujetos.

Para comprobar si este umbral es aplicable a la base de datos de CIPIC y para ganar más conocimiento respecto a la validez de los resultados obtenidos, se calcularon algunos valores estadísticos de la base de datos de CIPIC, y luego fueron comparados con los valores de distorsión obtenidos con los tres métodos de personalización aquí presentados. El objetivo principal de realizar este cálculo estadístico fue comparar las HRTF personalizadas con dos tipos de resultados: a) la selección aleatoria, es decir, una HRTF de la base de datos seleccionada al azar, y b) una HRTF genérica.

Selección aleatoria

Se calcularon las distorsiones (SD y CSD) entre las HRTF de todos los pares posibles de sujetos dentro de la base de datos; con N sujetos, las combinaciones son $(N \times N - N)/2$. Con los 35 sujetos de la base de datos, las combinaciones posibles son 595. Los 595 valores de distorsión obtenidos se promediaron; se muestran como *promedio entre pares de CIPIC* en el número 4 de los cuadros 6.1 para la SD y 6.2 para la CSD. Este valor representa la distancia promedio entre dos sujetos cualesquiera de la base de datos; pero también se puede entender como la distorsión que resultaría si la “personalización” se hiciera simplemente seleccionando un sujeto al azar de la base de datos.

Los resultados muestran que la red neuronal se desempeña peor, la regresión lineal múltiple producen valores de distorsión compaginados con los de la selección aleatoria, mientras que el método de mejor coincidencia antropométrica es ligeramente mejor. Es decir, uno de los métodos produce HRTF que son, consistentemente, mejores que seleccionar una HRTF al azar.

HRTF genérica

En las aplicaciones actuales de tecnología binaural se usan generalmente HRTF genéricas, por ejemplo las HRTF medidas sobre el maniquí acústico KEMAR [12, 29]. La base de datos de CIPIC contiene grabaciones de HRTF del maniquí KEMAR, usando dos modelos de oreja (pinna): grande y pequeña. Se presentan los resultados obtenidos con la oreja pequeña, pues con ésta se obtienen valores menores de error.

Se calculó la distorsión entre la HRTF del maniquí y cada uno de los sujetos de la base de datos, y los resultados se promediaron; se muestran en el número 5 de los cuadros 6.1 para la SD y 6.2 para la CSD. Nuevamente se observa que los valores de distorsión obtenidos con la red neuronal son mayores en todos los

casos. La regresión lineal múltiple es mejor solamente para el caso de la SD en la banda de 0-8 kHz, en todos los demás casos es peor o igual. El método de mejor coincidencia antropométrica tiene valores de SD menores que los del maniquí en tres de las cuatro bandas, pero en ninguna para el caso de la CSD.

Otro tipo de HRTF genérica es la HRTF que se obtiene al promediar las HRTF de un grupo de sujetos, en este caso, el promedio de las HRTF de todos los sujetos de la base de datos de CIPIC. Sin embargo esta HRTF no es buen candidato, pues la promediación elimina los picos y valles del espectro de la HRTF, que son las características de las HRTF que producen la sensación espacial, especialmente por encima de 5 kHz [15].

La HRTF promedio de la base de datos de CIPIC se calculó con el procedimiento presentado en la sección 4.2. Se aplicó la corrección de fase al espectro complejo (caso b), y se convirtió a escala logarítmica antes de hacer la promediación en ambos caso (a y b). Después de la promediación, se regresó el resultado a escala lineal, y se restituyó la fase original, para el caso (b).

Se calculó la distorsión entre la HRTF de cada sujeto y la HRTF promedio, y los resultados a su vez fueron promediados entre todos los sujetos, resultando en una métrica análoga al concepto de *desviación estándar* de la base de datos de CIPIC. El resultado se muestra con el número 6 de los cuadros 6.1 para la SD y 6.2 para la CSD.

La desviación espectral para este caso entrega los resultados con el menor error de entre todos los resultados presentados. Sin embargo, como ya se había mencionado, la HRTF promedio no es adecuada como HRTF en la práctica.

6.4. Efectos de los errores antropométricos en las HRTF

En la sección 3.2.2 se presentaron los errores obtenidos tras realizar la fotoantropometría automática (columnas 4 y 5 del cuadro 3.2) con respecto a las anotaciones del experto. En esta sección se analizan los efectos que esos errores introducen en las HRTF personalizadas. El análisis presentado en esta sección complementa otras investigaciones en curso, como la que se reporta en la referencia [216].

Se parte del vector $\tilde{\mathbf{x}}_p^t$ que contiene la información antropométrica de un sujeto objetivo t . Se utiliza el método de mejor coincidencia antropométrica (sección 4.1) para obtener las HRTF, $\tilde{\mathbf{h}}_F^t$, que corresponden con la antropometría $\tilde{\mathbf{x}}_p^t$.

El siguiente paso es introducir un error relativo arbitrario, e , dentro de uno de

los parámetros antropométricos del vector $\tilde{\mathbf{x}}_p^t$, es decir:

$$x_p \leftarrow e \cdot x_p \quad (6.1)$$

donde p es el índice del parámetro antropométrico siendo modificado. x_p se coloca nuevamente dentro del vector antropométrico al que, en esta forma modificada, denominamos $\tilde{\mathbf{x}}_p^e$. Usando el mismo procedimiento que para $\tilde{\mathbf{h}}_F^t$, se busca una HRTF que corresponda con la antropometría $\tilde{\mathbf{x}}_p^e$, resultando en el vector $\tilde{\mathbf{h}}_F^e$.

Se tienen entonces dos HRTF, una para la antropometría original, $\tilde{\mathbf{h}}_F^t$, y otra para la antropometría con un error introducido, $\tilde{\mathbf{h}}_F^e$. Sabremos cuanto error introduce el cambio de antropometría en las HRTF si se mide la distorsión (SD y CSD) entre $\tilde{\mathbf{h}}_F^t$ y $\tilde{\mathbf{h}}_F^e$.

El procedimiento se repite para todos los sujetos en la base de datos, y los resultados se promedian, obteniéndose así un estimado de distorsión producida (o introducida) por el error relativo e en el parámetro antropométrico p .

En el cuadro 6.3 se muestran los parámetros antropométricos con los que se trabajó (segunda columna) y sus respectivos errores relativos obtenidos en el procedimiento de foto-antropometría (tercera columna²). En la cuarta columna se muestra la distorsión espectral (SD) producida en las HRTF por cada error relativo de la tercera columna, y en la quinta columna se muestra la distorsión de espectro complejo (CSD); en paréntesis se muestra la desviación estándar para cada caso. Nuevamente los parámetros x_5 y x_7 no fueron considerados.³

Los valores de distorsión obtenidos son generalmente bajos, muchos de ellos son cero, lo que indica que los errores antropométricos promedio obtenidos con el método de fotoantropometría son suficientemente bajos, permitiendo aún así obtener la mejor coincidencia antropométrica con la base de datos de CIPIC.

En lo que respecta los otros errores, el valor de distorsión espectral más alto es 0.53 dB para d_4 , que es el parámetro con el cuarto error relativo más alto. También sucede que parámetros con errores relativos significativos no introducen error en la HRTF, como es el caso para x_4 que tiene el tercer error relativo más alto. Para el caso de la distorsión de espectro complejo, el error más alto es de 0.65 dB, para d_4 nuevamente. Todos los demás errores están por debajo de esta cantidad, y vuelven a aparecer casos en los que un error relativo significativo no introduce error en las HRTF. Evaluando en forma cualitativa se puede decir que los resultados obtenidos con SD y CSD son iguales.

Concluimos que algunos errores de fotoantropometría deben reducirse, específicamente aquellos que resultan en errores mayores a cero en las HRTF personalizadas; estos son, en orden descendente de relevancia, d_4 , d_7 , d_1 , x_2 , x_1 , d_3 , y d_6 .

²Son los mismos valores presentados anteriormente en el cuadro 3.2.

³Ver sección 3.2.2.

Modelo	Parámetro	Error relativo promedio	SD promedio dB (dB)	CSD promedio dB (dB)
Frontal	x_1	4.41 %	0.36 (1.50)	0.44 (1.85)
	x_2	2.86 %	0.37 (1.54)	0.37 (1.54)
	x_6	5.27 %	0 (0.00)	0 (0.00)
	x_7	109.91 %	—	—
	x_9	2.34 %	0 (0.00)	0 (0.00)
	x_{10}	8.36 %	0 (0.00)	0 (0.00)
	x_{12}	2.88 %	0 (0.00)	0 (0.00)
Lateral	x_3	2.20 %	0 (0.00)	0 (0.00)
	x_4	15.08 %	0 (0.00)	0 (0.00)
	x_5	128.19 %	—	—
	x_8	2.59 %	0 (0.00)	0 (0.00)
	x_{11}	2.20 %	0 (0.00)	0 (0.00)
Oreja	d_1	6.59 %	0.37 (1.54)	0.37 (1.54)
	d_2	17.83 %	0 (0.00)	0 (0.00)
	d_3	7.74 %	0.20 (1.16)	0.26 (1.56)
	d_4	12.82 %	0.53 (1.77)	0.65 (2.20)
	d_5	1.12 %	0 (0.00)	0 (0.00)
	d_6	4.09 %	0.17 (0.99)	0.18 (1.05)
	d_7	23.75 %	0.40 (1.65)	0.46 (1.91)

Cuadro 6.3: Errores relativos de photo-antropometría y sus efectos en las HRTF personalizadas, cuantificados con SD y CSD. La desviación estándar se muestra en paréntesis.

Capítulo 7

Personalización de HRTF a partir de la antropometría

Finalmente, para validar el procedimiento de personalización de HRTF a partir de la antropometría, se realizó la personalización de las HRTF de 15 sujetos disponibles. El método de personalización seleccionado fue el de mejor coincidencia antropométrica, descrito en la sección 4.1.

A todos los sujetos disponibles se les midieron sus HRTF dentro de la cámara anecoica, con conductos auditivos bloqueados, tal y como se describen las mediciones de la base de datos de CIPIC [20]. Sin embargo, no fue posible asegurar compatibilidad absoluta en la medición con respecto a la de CIPIC, esto debido a la ausencia de algunos detalles como la ubicación exacta de la sonda del micrófono sobre la *pinna*, etc.

En la figura 7.1 se muestra la magnitud de las HRTF medidas sobre el oído derecho y con la fuente en la posición frontal (ángulo $\theta = \phi = 0^\circ$) de tres de los 15 sujetos empleados. Cada HRTF está comparada con la HRTF que resulta de la personalización a través del método de mejor coincidencia antropométrica, usando la base de datos de CIPIC. Se puede apreciar en lo general una buena correspondencia entre la HRTF medida y la personalizada, especialmente para el caso de los sujetos 11 y 14, en donde se observa coincidencia entre varios rasgos espectrales (crestas y valles).

La distorsión espectral medida entre las HRTF originales y sus HRTF personalizadas se muestra en el cuadro 7.1. Los valores de distorsión se muestran en particular para los tres sujetos presentados en la figura 7.1; también se muestran los valores promedio y desviación estándar para los 15 sujetos a quienes se les personalizaron sus HRTF.

Los valores de distorsión espectral son menores en la banda de 0–3.4 kHz, y

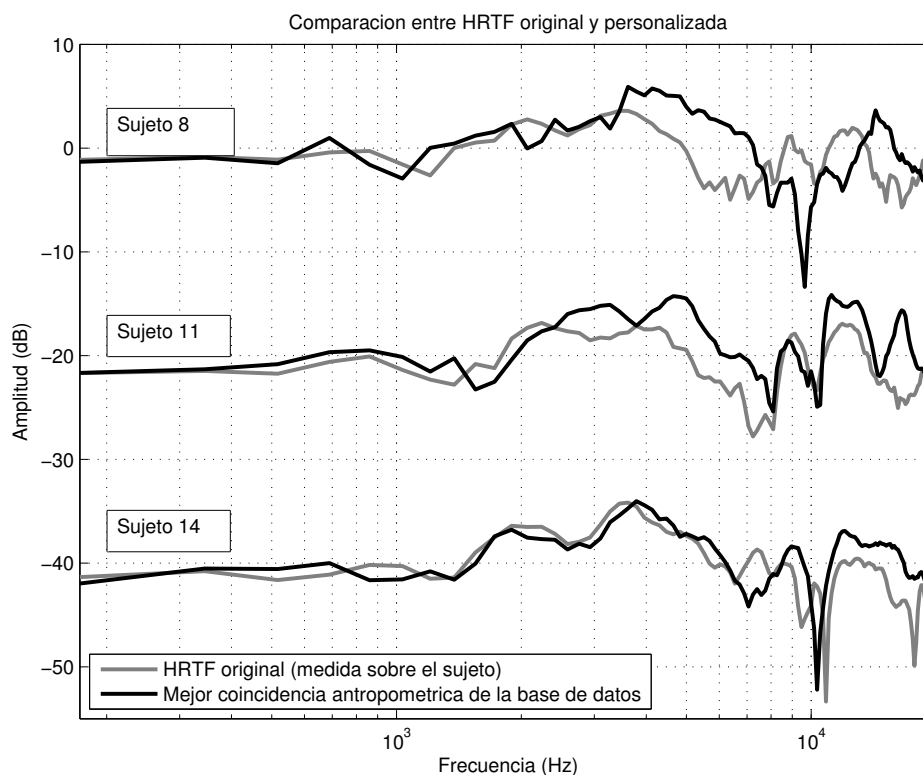


Figura 7.1: Magnitud de la HRTF medida para tres sujetos, comparadas con las HRTF personalizadas con el método de mejor coincidencia antropométrica de la base de datos de CIPIC. Las curvas se desplazan en pasos de -20 dB con propósito de claridad.

se incrementan con la frecuencia. Para el caso del sujeto 8, la distorsión espectral es la mayor de todas para las bandas de 0–8 kHz, 0–17 kHz, y 0–22 kHz, lo cual concuerda con las discrepancias entre las HRTF original y personalizada, observadas en la figura 7.1.

Sin embargo, aún con la incertidumbre de la compatibilidad entre las mediciones de CIPIC y las aquí presentadas, los niveles de distorsión son, en lo general, bajos.

	SD (dB)				CSD (dB)			
	Límite superior (kHz)				Límite superior (kHz)			
	3.4	8	17	22	3.4	8	17	22
Sujeto 8	2.5	7.1	7.9	8.0	6.5	6.5	8.1	8.0
Sujeto 11	3.9	6.4	6.3	6.5	4.9	4.7	5.2	5.2
Sujeto 14	3.1	4.3	6.9	7.9	2.8	5.2	5.7	5.8
Promedio	3.6	5.5	7.0	7.5	4.2	6.0	6.5	6.5
Desv. estándar	1.0	1.1	0.7	0.7	1.1	1.4	1.2	1.18

Cuadro 7.1: Valores de distorsión espectral SD y CSD (en decibeles) entre HRTF medidas y HRTF personalizadas.

Capítulo 8

Relevancia de los parámetros antropométricos

Resulta evidente la necesidad de analizar la relevancia que tiene cada uno de los parámetros antropométricos en las HRTF. Este capítulo aborda dicho tema, y los resultados obtenidos permitirán averiguar si los parámetros incluidos en la base de datos de CIPIC son, efectivamente, adecuados para la personalización de HRTF, o si será necesario proponer en trabajos futuros una nueva colección de parámetros antropométricos que sean acústicamente más relevantes.

El análisis presentado en este capítulo tiene dos objetivos principales: 1) analizar la relevancia de cada parámetro antropométrico en la personalización de las HRTF, en función del ángulo y la frecuencia, y 2) comparar los resultados presentados en la bibliografía [22, 205, 209, 210, 213, 217].

El problema se abordó como una extensión del método de mejor coincidencia antropométrica; la hipótesis fundamental es que cuando dos personas tienen antropometría similar, sus HRTF serán compatibles. La compatibilidad se cuantifica como el error o distancia euclideana entre los dos vectores normalizados de antropometría.

Una solución simple sería seleccionar un grupo reducido de parámetros antropométricos de entre los disponibles en la base de datos, hacer la personalización de las HRTF, y medir y almacenar el error (distorsión espectral). El siguiente paso consiste en agregar un parámetro antropométrico adicional al grupo, se hace una nueva personalización usando ahora el nuevo grupo de parámetros antropométricos y se mide nuevamente el error; si este error es mayor que el que se obtuvo originalmente, significa que el parámetro antropométrico que se agregó repercute en la personalización, y por lo tanto, el parámetro no debe ser considerado. Por otro lado, si el error obtenido en la nueva personalización es menor que el error

original, significa que el nuevo parámetro es necesario en la personalización.

Se puede repetir el procedimiento evaluando todos los grupos o combinaciones posibles de parámetros antropométricos, buscando las combinaciones que resulten en el mínimo error de personalización. El procedimiento puede extenderse de forma tal que sea posible obtener una medida de la relevancia porcentual de cada parámetro antropométrico en función del ángulo de la HRTF y la frecuencia. Dicho método se describe a continuación.

8.1. Método

Se trabajó con la base de datos de CIPIC, que contiene la antropometría y las HRTF de $s = 35$ sujetos. El primer paso es obtener una matriz S_H en donde cada uno de sus elementos (m, n) sea el error de distorsión espectral (SD) entre las HRTF del sujeto m y el sujeto n (esta matriz es análoga a la matriz de correlación de las HRTF).

Sea $p = 27$ el número total de parámetros antropométricos disponibles en la base de datos, el total de combinaciones (o subgrupos) de parámetros antropométricos posibles es $c = 2^p = 2^{27} = 1.34 \times 10^8$. Se programa entonces un ciclo para evaluar cada combinación o subgrupo c . Para cada combinación se obtiene una matriz S_A^c en la que cada uno de sus elementos (m, n) es el la distancia (error) entre los vectores normalizados¹ del subgrupo c de parámetros antropométricos de los sujetos m y n . De igual forma, esta matriz es un análogo de la matriz de correlación de la antropometría de los sujetos.

El escenario ideal es aquel en el que las matrices S_H y S_A son cualitativamente lo más similar posibles. Y debido a que se están evaluando todas las combinaciones, es posible almacenar los mejores resultados de forma tal que al final del procedimiento se tenga acumulado un porcentaje de presencia de cada parámetro antropométrico en los mejores resultados, obteniendo así una medida efectiva de la relevancia de cada parámetro antropométrico.

Algoritmo

A continuación se presenta el algoritmo en *pseudo-código*. Las expresiones que usan matrices respetan la sintaxis de MATLAB:

```
s = numero de sujetos
H = [matriz de HRTF (sujeto, frecuencia)]
X = [matriz de Antropometría (sujeto, parámetro_antropométrico)]
angulos = [vector de angulos]
```

¹La distancia y la normalización se calculan con el procedimiento presentado en la sección 4.1.

```

frecuencias = [vector de frecuencias]
for every angulo in angulos do
  for every frecuencia in frecuencias do
    H' = H[:,frecuencia]           ▷ (subset of H specific for this frequency)
    for m,n from 1 to s do
      SH(m,n) = SD(H'(m,:),H'(n,:))   ▷ (matriz de SD entre las HRTF)
    end for
    for PA = every case from 1 to 2p do ▷ (para cada combinación de PA)
      X' = X[:,PA]                   ▷ (subset of X specific for this case)
      l = length(PA)                 ▷ (número de columnas de X')
      X' =  $\frac{X' - \bar{X}'}{std(X')}$            ▷ (normalizar X')
      for m,n from 1 to s do
        SA(m,n) = (|X'(m,:) - X'(n,:)|) / l
                                     ▷ (error promedio entre antropometrías)
      end for
      fc = frobenius(SH, SA)           ▷ (módulo de Frobenius)2
      if fc < máximo(resultados anteriores) then
        almacena PA en la posición del máximo en resultados anteriores
                                     ▷ (almacenar los 100 mejores casos...
                                     ▷ ... para este ángulo y frecuencia)
      end if
    end for
  end for
end for

```

Este algoritmo es, sin embargo, muy costoso en términos computacionales, pues requiere la evaluación de todas las 2^p combinaciones posibles de parámetros antropométricos a distintos ángulos y para distintas frecuencias; se decidió entonces limitar el estudio dentro de los criterios que se exponen a continuación:

Ángulos Los ángulos que fueron empleados para recopilar la base de datos de CIPIC no son equidistantes, por ejemplo, entre 0° y 90° los ángulos disponibles son: 0° , 5° , 10° , 15° , 20° , 25° , 30° , 35° , 40° , 45° , 55° , 65° y 80° , es importante notar que el ángulo de 90° no está disponible.

Los ángulos elegidos para este análisis, mostrados en la figura 8.1, son aquellos que se encuentran, principalmente, en la zona ipsilateral al oído derecho, que es el oído seleccionado para el análisis.

²El módulo de Frobenius cuantifica la diferencia cualitativa de dos matrices de las mismas dimensiones, es cero cuando ambas matrices son iguales.

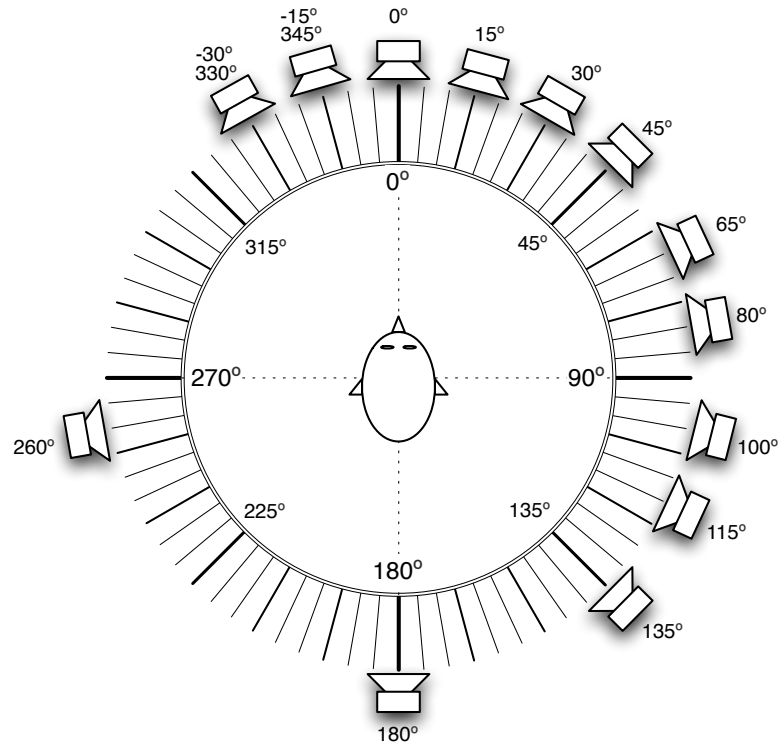


Figura 8.1: Esquema de los ángulos de procedencia del sonido empleados en el estudio. El ángulo de 0° corresponde a la dirección frontal, 90° corresponde a la oreja derecha del sujeto.

Frecuencias Se trabajó con las componentes de frecuencia de los espectros de HRTF desde 172 Hz hasta 3.96 kHz.

Parámetros antropométricos La base de datos de CIPIC incluye información de 17 parámetros antropométricos de la cabeza y el torso de las personas y 10 parámetros antropométricos para caracterizar cada *pinna* de los sujetos. Estos 27 parámetros antropométricos se listan en el cuadro 8.1.

Se tienen entonces 2^{27} combinaciones posibles, pero debido a que la cantidad de tiempo necesario para calcular todas las combinaciones resultó prohibitivo, se descartaron dos parámetros: x_{14} y x_{15} . La altura del sujeto, x_{14} , se descarta fácilmente ya que las mediciones se hicieron con el sujeto sentado; y x_{15} (altura del sujeto sentado) solamente podría influir en las reflexiones que produzcan las piernas del sujeto, pero éstas aparecerían como las últimas componentes temporales de las HRIR, que invariablemente tienen amplitud despreciable. De esta forma, el problema se simplificó a 2^{25} combinaciones posibles.

Cabeza, cuello y torso			Oreja		
No.		Parámetro	No.		Parámetro
1	x1	head width	18	d1	cavum concha height
2	x2	head height	19	d2	cymba concha height
3	x3	head depth	20	d3	cavum concha width
4	x4	pinna offset down	21	d4	fossa height
5	x5	pinna offset back	22	d5	pinna height
6	x6	neck width	23	d6	pinna width
7	x7	neck height	24	d7	intertragal incisure width
8	x8	neck depth	25	d8	cavum concha depth
9	x9	torso top width	26	$\theta 1$	pinna rotation angle
10	x10	torso top height	27	$\theta 2$	pinna flare angle
11	x11	torso top depth			
12	x12	shoulder width			
13	x13	head offset forward			
14	x14	height			
15	x15	seated height			
16	x16	head circumference			
17	x17	shoulder circumference			

Cuadro 8.1: Lista completa de los parámetros antropométricos incluidos en la base de datos de CIPIC [20].

Tiempo de cálculo

El desempeño del programa se analizó con la herramienta *Profiler* de MATLAB, con la que es posible observar el porcentaje de ciclos de procesador que se requieren para ejecutar cada línea del programa. Se identificaron las rutinas más lentas y se optimizó el código para disminuir el tiempo de cálculo dentro de lo posible. Algunas decisiones de diseño del algoritmo se tomaron en torno a las observaciones de esta herramienta, por ejemplo, la decisión de conservar solamente los cien mejores resultados del cálculo se tomó tras entender que el tiempo y memoria requerida para el cálculo aumenta en función del tamaño de los resultados almacenados. Se vuelve prohibitivo intentar almacenar más resultados.

Una vez que se tuvo un código eficiente, se hicieron pruebas preliminares, midiendo el tiempo de cálculo para distinto número de parámetros antropométricos y se extrapoló para la colección completa de 27 parámetros, obteniendo un estimado de 150 horas de cálculo en un procesador 2 GHz, de un solo núcleo. Fue por esta razón que se decidió reducir el número de parámetros a 25, lo que implica un cálculo de alrededor de 37 horas.

Cada cálculo corresponde a una combinación de (ángulo, frecuencia); se tra-

bajó con 23 frecuencias y 13 ángulos, se presenta entonces el resultado de 11,063 horas (15 meses) de procesamiento. El cálculo se repartió entre varias computadoras, de las cuales se tuvieron hasta seis operando en paralelo.

8.2. Resultados

Para comenzar con el análisis de los resultados, en la figura 8.2 se muestran los resultados del parámetro antropométrico $d3$ (*cavum concha width*). La figura contiene tres gráficas. La gráfica de la izquierda indica cuál es el porcentaje de relevancia de este parámetro en función del ángulo y la frecuencia. Para cada combinación de (ángulo, frecuencia) sobre la figura se observa un tono de gris, que corresponde con la escala de relevancia porcentual (%) que se muestra a la izquierda de la figura; zonas claras (blanco) corresponden a 100% de relevancia y zonas oscuras (negro) implican irrelevancia.

En la gráfica se observa que para 0° y 0.17 kHz la relevancia es de 100%, esto significa que para esta combinación, este parámetro antropométrico tuvo presencia en los cien mejores resultados que se encontraron. Por el contrario, para el caso de 0° y 3.96 kHz, donde la relevancia es muy cercana a 0%, este parámetro antropométrico casi no tuvo presencia en el grupo de los mejores resultados para dicha combinación.

En lo general se observan zonas o cúmulos de relevancia (zonas claras) y zonas de irrelevancia. Sin embargo el análisis de estos resultados no es sencillo, el comportamiento incluso pareciera ser aleatorio en algunas zonas (aunque no lo es, como se comprobará más adelante); es por esta razón que se presentan dos gráficas adicionales dentro de la misma figura; donde se analiza la misma información pero con otra perspectiva:

La gráfica de la esquina superior derecha muestra la relevancia porcentual promedio en función del ángulo. Es el resultado de promediar, para cada ángulo, los valores de relevancia de todas las frecuencias (desde 0.17 hasta 3.96 kHz). En la gráfica también se muestra una línea horizontal sólida que representa el valor promedio; le acompañan dos líneas punteadas, que representan una desviación estándar por encima y por debajo del promedio. Se observa que, en lo general, la relevancia de este parámetro antropométrico varía poco en función del ángulo.

La gráfica de la esquina inferior derecha muestra el resultado de promediar, para cada frecuencia, los valores de relevancia de todos los ángulos; esto permite analizar la relevancia porcentual promedio en función de la frecuencia. De igual forma, las líneas sólidas y punteadas indican el promedio y la desviación estándar, respectivamente. En este caso se observan grandes variaciones de relevancia promedio al variar la frecuencia. En las bajas frecuencias la relevancia es alta, sobrepasa incluso 90%; al aumentar la frecuencia, la relevancia disminuye

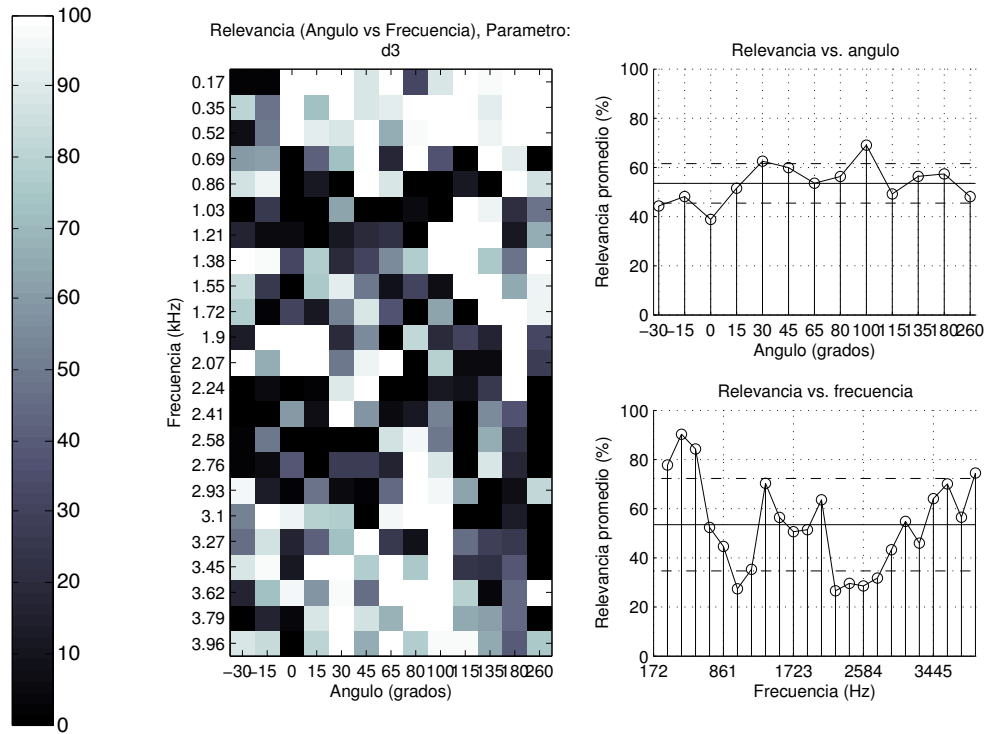


Figura 8.2: Gráficas de la relevancia del parámetro antropométrico $d3$ (*cavum concha width*) en función del ángulo y la frecuencia.

hasta llegar a un mínimo local en 1.03 kHz. En frecuencias mayores la relevancia aumenta y disminuye, hasta llegar a otro mínimo en 2.24 kHz, y de ahí aumenta al subir la frecuencia hasta un máximo local en 3.96 kHz.

Se presenta a continuación el resultado obtenido para otro parámetro antropométrico cuyo comportamiento es contrastante, se trata del parámetro x_{12} (*ancho del hombro*), en la figura 8.3.

En este caso, la relevancia es en lo general cercana a 0%, en función tanto de la frecuencia como del ángulo. Este parámetro solamente es relevante (cerca del 100%) en la dirección ipsilateral (cerca a 80°) y a 1 kHz; para la mayoría de los demás casos es irrelevante. El comportamiento de la relevancia de este parámetro antropométrico es claramente distinto al del caso anterior, y se descarta así que los resultados obtenidos sean aleatorios.

Se presentan finalmente los resultados para los 25 parámetros antropométricos estudiados, pero por brevedad se evita hacer un análisis minucioso de cada uno. En las figuras 8.4, 8.5 y 8.6 se presentan las gráficas que corresponden a los parámetros antropométricos que caracterizan a la cabeza y torso del sujeto, es

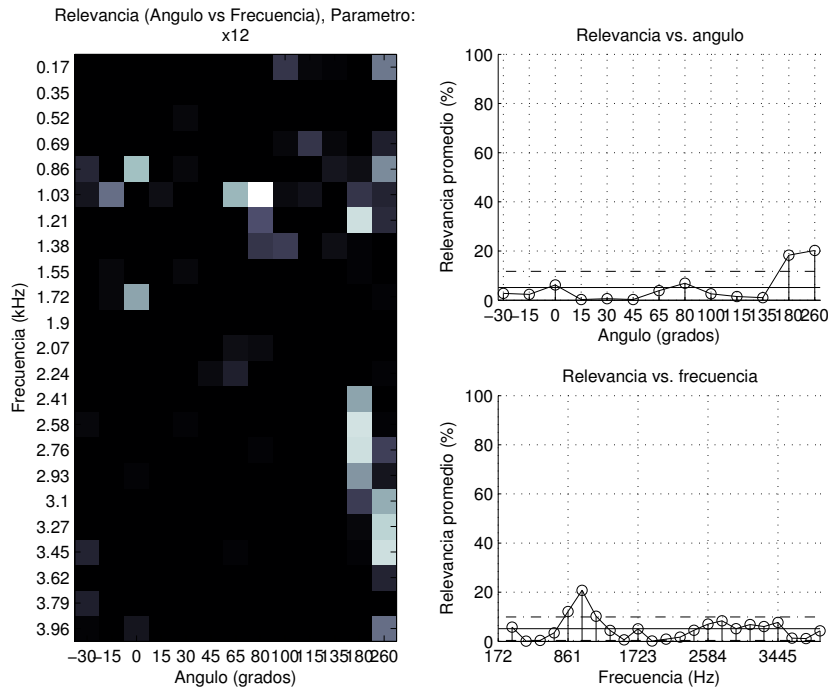


Figura 8.3: Gráfica de la relevancia del parámetro antropométrico x_{12} (*ancho del hombro*) en función del ángulo y la frecuencia.

decir, de x_1 a x_{13} , x_{16} y x_{17} , y en las figuras 8.7 y 8.8 se presentan los resultados para los parámetros de la *pinna*, es decir, de d_1 a d_8 , θ_1 y θ_2 .

Comparación entre parámetros antropométricos

Aunque cada gráfica presentada permite estudiar el comportamiento individual de los parámetros antropométricos, queda pendiente hacer una comparación entre los mismos. Se presenta entonces la métrica de la relevancia porcentual promedio, que se obtiene al promediar todos los valores de relevancia (de todos los ángulos y todas las frecuencias) de cada parámetro. Esta medida permite observar finalmente qué parámetros antropométricos de la base de datos de CIPIC tienen mayor relevancia.

El resultado de esta comparación se muestra como barras en la figura 8.9 (pag. 63). La altura de cada barra indica el valor de la relevancia porcentual promedio de cada parámetro. Estos se muestran ordenados de mayor a menor relevancia. Así, claramente, el parámetro antropométrico más relevante de la base de datos de CIPIC es d_3 (*cavum concha width*) con 53.5 % de relevancia, y el menos relevante de todos es x_{12} (*shoulder width*) con 5.14 %. Estos son los mismo parámetros cuyos resultados ya fueron analizados y discutidos en párrafos anteriores y en las figuras

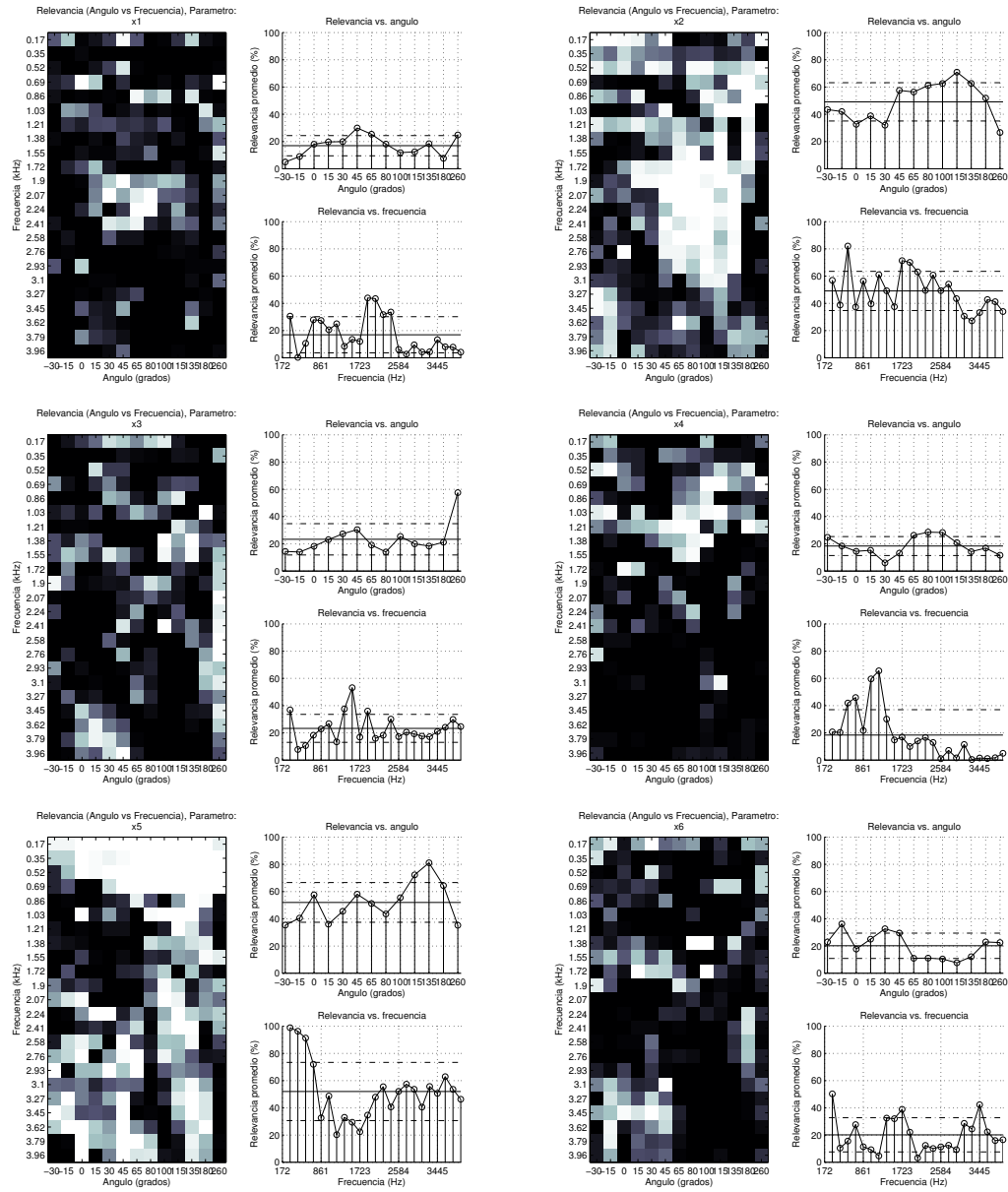


Figura 8.4: Resultados de relevancia para los parámetros antropométricos x1 - x6.

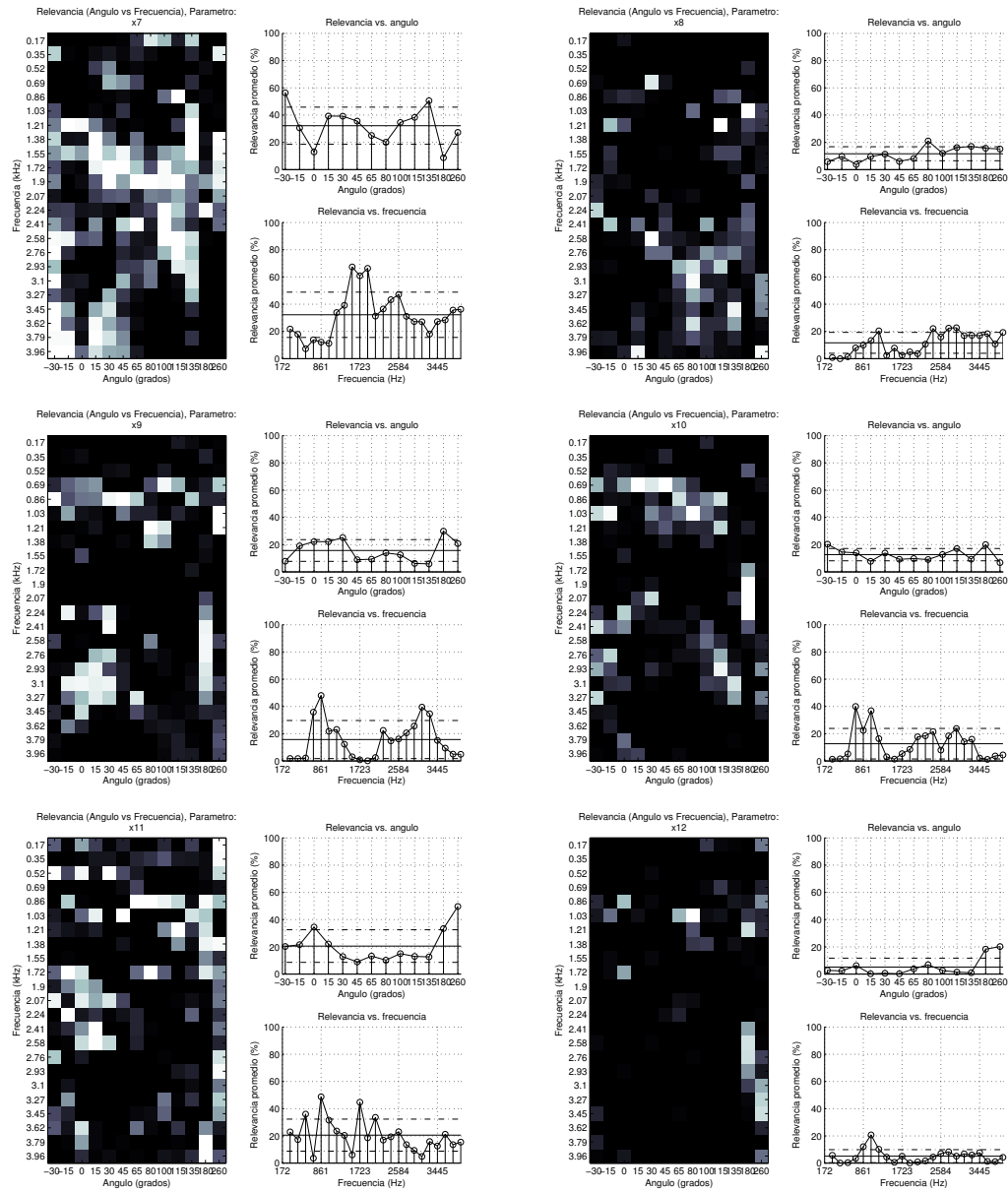


Figura 8.5: Resultados de relevancia para los parámetros antropométricos x7 - x12.

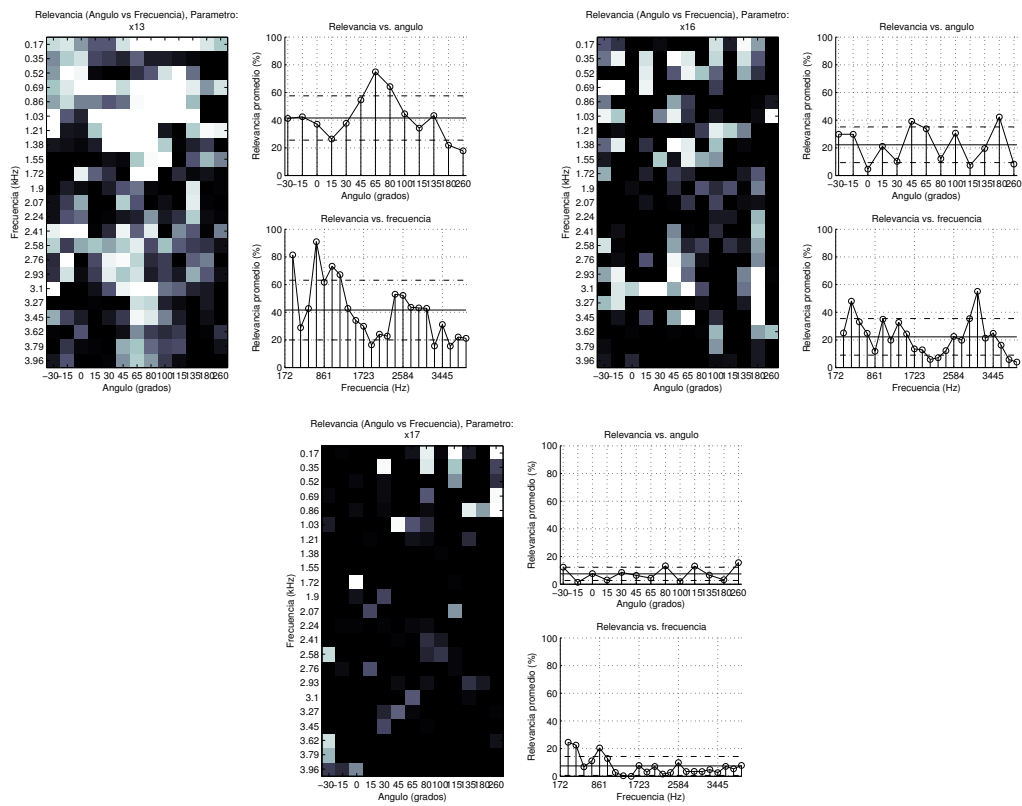


Figura 8.6: Resultados de relevancia para los parámetros antropométricos x13, x16 y x17.

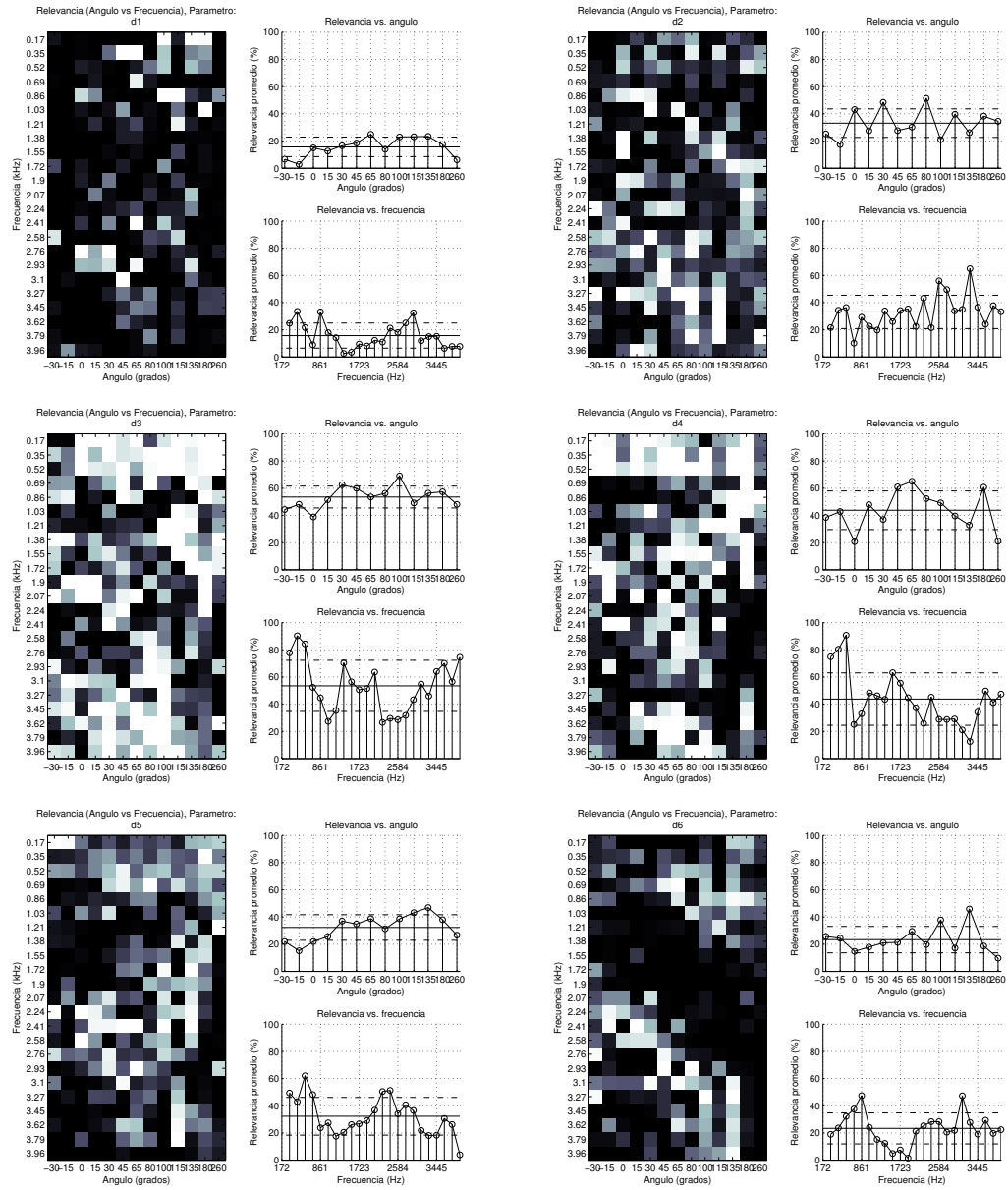


Figura 8.7: Resultados de relevancia para los parámetros antropométricos de la oreja, d1 - d6.

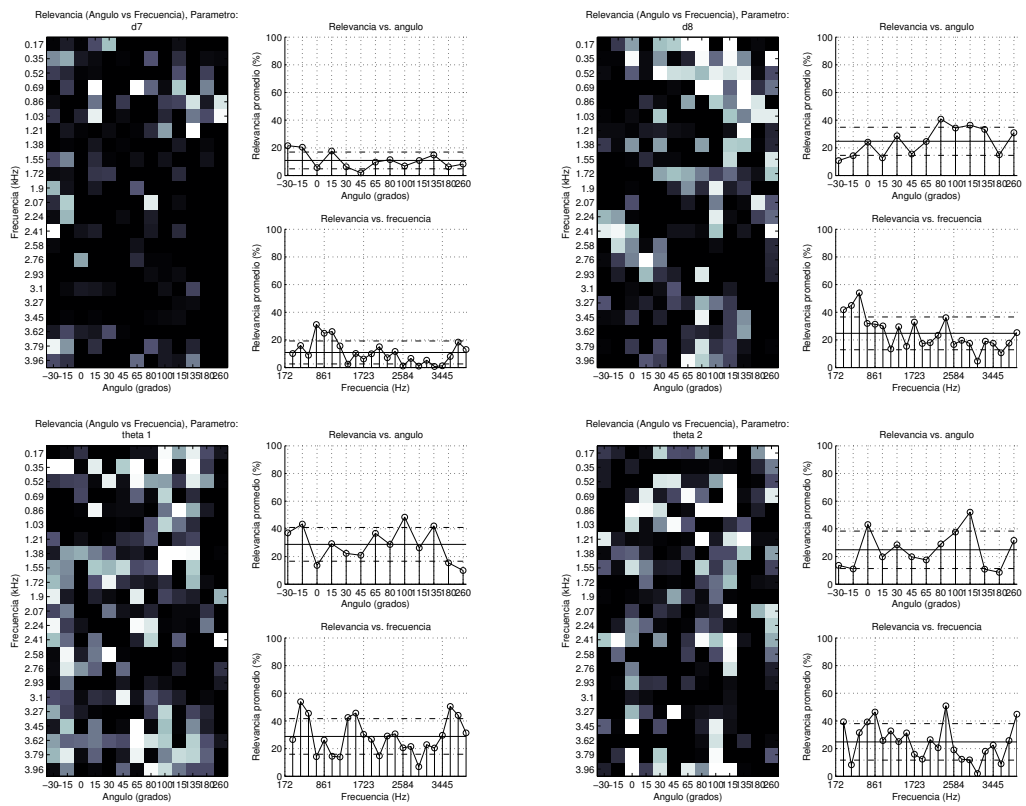


Figura 8.8: Resultados de relevancia para los parámetros antropométricos de la oreja, d_7 , d_8 , θ_1 y θ_2 .

8.2 y 8.3, respectivamente.

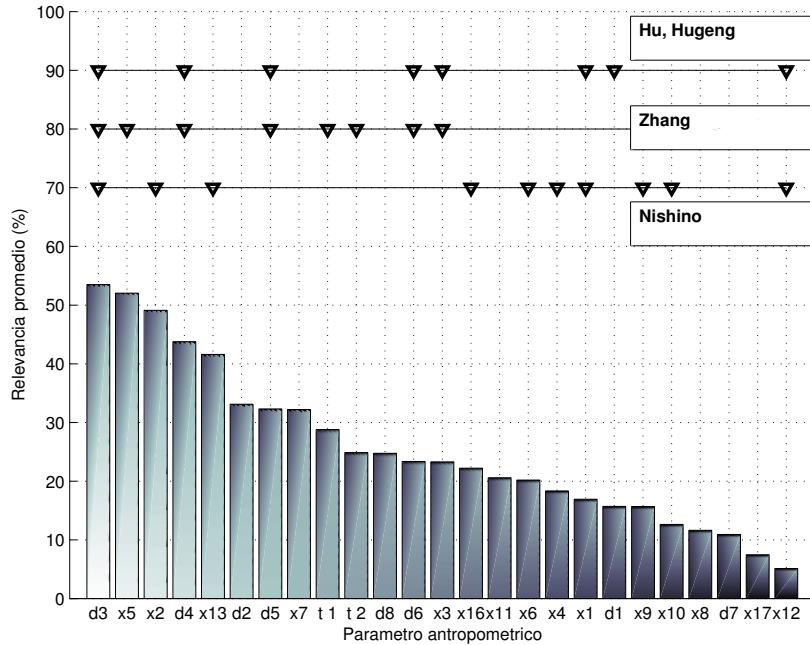


Figura 8.9: Gráfica de la relevancia promedio de todos los parámetros antropométricos de la base de datos de CIPIC, ordenados de mayor a menor. Se agregan los parámetros antropométricos empleados por distintos autores (Hu, Hugeng, Zhang y Nishino).

El listado de los parámetros antropométricos, reordenado de acuerdo a estos resultados de relevancia porcentual promedio se muestra en el cuadro 8.2 (pag. 64).

8.3. Comparación con la bibliografía

Los trabajos reportados en la bibliografía en los que se estudia la relevancia de los parámetros antropométricos [22, 205, 209, 210, 213, 217] hacen el cálculo de la correlación entre los parámetros y las HRTF; y sin embargo, al final seleccionan arbitrariamente un grupo de parámetros, siguiendo algún otro criterio, para después hacer la personalización de las HRTF. En la figura 8.9 también se indican, con flechas, estos parámetros antropométricos seleccionados por otros autores.

Hu [205, 213] y Hugeng [209, 210, 217] trabajan con regresión lineal múltiple entre la antropometría y las HRTF. Seleccionan ocho parámetros antropométricos de forma arbitraria (se muestran en la primera fila de la figura 8.9). Se observa que

Parámetro	Modelo	Relevancia
cavum concha width	d3 oreja	53.5 %
pinna offset back	x5 lateral	52.0 %
head height	x2 frontal	49.1 %
fossa height	d4 oreja	43.8 %
head offset forward	x13	41.6 %
cymba concha height	d2 oreja	33.1 %
pinna height	d5 oreja	32.3 %
neck height	x7 frontal	32.2 %
pinna rotation angle	t1	28.8 %
pinna flare angle	t2	24.9 %
cavum concha depth	d8	24.8 %
pinna width	d6 oreja	23.4 %
head depth	x3 lateral	23.3 %
head circumference	x16	22.2 %
torso top depth	x11 lateral	20.6 %
neck width	x6 frontal	20.2 %
pinna offset down	x4 lateral	18.3 %
head width	x1 frontal	16.9 %
cavum concha height	d1 oreja	15.7 %
torso top width	x9 frontal	15.7 %
torso top height	x10 frontal	12.6 %
neck depth	x8 lateral	11.6 %
intertragal incisure width	d7 oreja	10.9 %
shoulder circumference	x17	7.47 %
shoulder width	x12 frontal	5.14 %

Cuadro 8.2: Relevancia porcentual promedio de cada parámetro antropométrico. El valor de la relevancia promedio (%) se presenta en la columna de la izquierda.

estos autores seleccionaron parámetros que, de acuerdo a nuestro resultado, son tanto relevantes como irrelevantes. Es decir, la selección arbitraria que realizan no es acertada si ésta se evalúa con los resultados obtenidos en esta investigación (una selección acertada implicaría que los parámetros elegidos aparezcan del lado izquierdo de la gráfica, que corresponden con los parámetros de mayor relevancia).

Nishino [22] elige nueve parámetros de forma arbitraria, sin embargo, no trabaja con la base de datos de CIPIC, sino una propia, la de CIAIR [22]. Algunos parámetros tienen un equivalente en la base de datos de CIPIC y otros pueden inferirse, siendo así, las equivalencias (10 parámetros) se indican en la tercera fila de la figura 8.9. Se observa que, al igual que Hu o Hugeng, la selección arbitraria de Nishino resulta en parámetros que no necesariamente son relevantes, de hecho, la mayoría de los parámetros se encuentran del lado derecho de la figura, que corresponde a los parámetros menos relevantes.

Finalmente, Zhang [218] realiza un estudio estadístico de la relevancia de cada parámetro antropométrico usando el análisis de componentes principales; selecciona ocho parámetros (se muestran en la segunda fila de la figura 8.9). Al contrario de los casos anteriores, esta selección de parámetros no es arbitraria, sino objetiva. Se observa que los ocho parámetros que elige se encuentran del lado izquierdo de la gráfica, que corresponde a la zona de los parámetros más relevantes, es decir, el análisis presentado por Zhang, y el análisis presentado en este capítulo, en cierta forma, se validan mutuamente.

Capítulo 9

Discusión

En los resultados presentados en los capítulos anteriores, se observó que las HRTF pueden personalizarse, hasta cierto grado, a partir de la antropometría y la base de datos de CIPIC. El método de selección por mejor coincidencia antropométrica produce los valores más bajos de distorsión espectral de entre los tres métodos evaluados. Su desempeño promedio es, en todos los casos, mejor que la selección aleatoria de una HRTF de entre la base de datos, y en algunos casos mejor que usar la HRTF genérica del maniquí KEMAR, especialmente si el desempeño se evalúa con la métrica de la distorsión espectral.

Sin embargo, los resultados obtenidos con los métodos de personalización son muy cercanos a los métodos genéricos, por lo que es necesario continuar la investigación en busca de formas de mejorar el resultado. Los parámetros estadísticos obtenidos de la base de datos de CIPIC (promedio entre pares, HRTF del maniquí y HRTF promedio) deben considerarse en investigación futura de esta área como valores de referencia para evaluar los resultados de los métodos de personalización de HRTF. Al menos así debería hacerse cuando no es posible hacer pruebas subjetivas, como fue el caso en esta investigación por no haber tenido acceso a los sujetos de la base de datos de CIPIC para validación.

Los temas que deben abordarse en investigación futura son la selección de parámetros antropométricos que sean suficientes y acústicamente relevantes, el número y selección de sujetos para conformar una base de datos estadísticamente más representativa de una población¹, y el desarrollo de métodos mejorados o nuevos métodos de personalización de HRTF. Algunas contribuciones relevantes en estos temas han sido presentados en [216, 218, 237, 238].

¹La base de datos de CIPIC fue conformada principalmente por estudiantes universitarios de la Universidad de California en Davis

Relevancia de los parámetros antropométricos

De los resultados reportados en el capítulo 8 se puede concluir lo siguiente: es posible analizar y cuantificar la relevancia de cada parámetro antropométrico en las HRTF en función de la frecuencia y el ángulo; claramente, el comportamiento de cada parámetro varía en función de estas variables.

Fue posible comparar los parámetros antropométricos empleados en la bibliografía. Los resultados obtenidos en este trabajo sugieren que aquellos parámetros que se escogieron en forma arbitraria, en otros trabajos, no integran el grupo más relevante posible de parámetros antropométricos. La selección arbitraria no es acertada. Por otro lado, aquellos parámetros seleccionados en forma objetiva y los resultados obtenidos en este trabajo son congruentes entre sí.

Queda pendiente analizar la posibilidad de desarrollar un método de personalización de HRTF a partir de aquellos parámetros antropométricos relevantes; sin embargo, debido a que la relevancia de cada parámetro es función tanto del ángulo como de la frecuencia, sería necesario un método con una regresión distinta para cada combinación de (ángulo, frecuencia).

Finalmente, debe resaltarse que en este trabajo no se reportan resultados para todos los ángulos y frecuencias. Si se decidiera extender este estudio, será necesario evaluar la utilidad de algún otro algoritmo de búsqueda que opere en menor tiempo; algunos algoritmos útiles son el algoritmo genético o el método de Montecarlo, que en su momento se descartaron por falta de tiempo para desarrollarlos y evaluarlos.

9.1. Contribuciones originales

A continuación se listan las aportaciones originales al campo de la personalización de HRTF a las que se llegó en este trabajo.

- Se propone y demuestra un método de personalización automática de HRTF a partir de fotografías del sujeto, constituye la primera vez que se reporta un método automático de personalización en la bibliografía.
- Se propone y demuestra el uso del algoritmo de modelos de forma activa (ASM) para obtener la fotoantropometría de una persona en forma automática.
- Se propone un nuevo método no lineal de personalización de HRTF a partir de la antropometría de la persona en la que se utilizan Redes Neuronales Artificiales.

-
- Se propone una nueva métrica para evaluar la distorsión entre dos espectros de naturaleza compleja, la métrica que se empleada anteriormente solamente toma en cuenta la magnitud, no la fase. La nueva métrica tiene aplicación en cualquier ámbito donde sea necesario caracterizar errores en espectros.
 - Se proponen por primera vez criterios de análisis estadístico de la base de datos de HRTF con los que es posible establecer niveles de error mínimos que deben alcanzar los métodos de personalización de HRTF.
 - Se propone y presenta un arreglo experimental para el retratado de sujetos, para utilizarse en tareas de fotoantropometría en trabajos futuros.
 - Se presenta un nuevo método de análisis y cuantificación de la relevancia de cada parámetro antropométrico de la base de datos y, por primera vez, una cuantificación porcentual de la relevancia de cada uno en las HRTF.

9.2. Recomendaciones

Se propone sustituir la métrica de la distorsión espectral por una medida de correlación entre HRTF; esto se debe a que la ubicación de las características espectrales de las HRTF es el rasgo que provee al usuario de la información necesaria para la percepción espacial, y la métrica de la distorsión espectral no cuantifica estas diferencias.

Se recomienda entonces proponer una métrica complementaria a la distorsión espectral que sea útil para evaluar el error de las HRTF personalizadas. Dicha métrica debe tomar en cuenta la correlación entre las dos HRTF.

9.3. Trabajo futuro

La investigación reportada en este trabajo puede continuarse en las direcciones siguientes:

- Es necesario reducir el error obtenido en la etapa del ajuste del modelo de forma activa sobre los bordes de la imagen. Este aspecto es importante particularmente en las zonas en las que se ubican los puntos somatométricos que definen los parámetros antropométricos en los que se obtuvieron los errores más altos.
 - Ante la dificultad de proponer una colección efectiva de parámetros antropométricos que sean acústicamente relevantes en las HRTF, se debe evaluar la idea de no utilizar la antropometría en lo absoluto en la personalización
-

de HRTF. Se propone entonces utilizar el modelo ajustado de la persona (sin necesidad de obtener la fotoantropometría) y usar las ponderaciones de las componentes principales que dan forma al modelo como precursores (a través de la red neuronal o la regresión) de la síntesis de la HRTF personalizada. Esto es en parte posible a raíz de la publicación ² de parte del material fotográfico empleado para la síntesis de la base de datos de CIPIC [20].

- La idea anterior puede extenderse al caso de los modelos tridimensionales del sujeto. La desventaja actual del uso de modelos tridimensionales es la gran cantidad de recursos necesarios para calcular las HRTF usando el BEM. Sin embargo, el modelo 3D podría proyectarse en una base ortogonal de componentes principales, cuyas ponderaciones podrían usarse para calcular las ponderaciones de las componentes principales de las HRTF y así sintetizar la HRTF del usuario. Este procedimiento evita el uso del BEM.
- Definir nuevas métricas de caracterización del error entre HRTF que tomen en cuenta variables interaurales, como la diferencia de tiempo interaural o la diferencia de nivel interaural.

² Disponible en la dirección: http://interface.cipic.ucdavis.edu/data/2012_Rev.zip [U.C.: 6 de junio, 2013].

Capítulo 10

Conclusiones

En este trabajo se usó en forma exitosa el algoritmo de visión computacional, modelos de forma activa, para obtener en forma automatizada los parámetros antropométricos de un sujeto humano a partir de retratos obtenidos en condiciones controladas. Se usó esta información antropométrica en distintos métodos de inferencia de HRTF personalizadas a partir de una base de datos. Dos de los métodos se retoman de la literatura: mejor coincidencia antropométrica, y regresión lineal múltiple, además de un intento utilizando redes neuronales artificiales. Los métodos fueron evaluados en términos de la distorsión espectral SD, y la distorsión de espectro complejo CSD, medidos entre la HRTF original contenida en la base de datos, y la HRTF personalizada. Se observó que en todos los casos, incluyendo resultados publicados previamente por otros autores, la distorsión espectral no se reduce por debajo de niveles aceptables; el método de mejor coincidencia antropométrica, que entrega el menor error, se desempeña dentro de ± 1.3 dB de la HRTF genérica, y dentro de ± 0.5 dB de la selección aleatoria. Se sugiere la posibilidad de que esto se deba a que la base de datos empleada (y otras disponibles como [22, 101, 105]) contenga información insuficiente en lo que respecta al número de sujetos que las conforman, o inadecuada en lo que respecta a los parámetros antropométricos seleccionados.

Para resolver estos detalles es necesario realizar investigación futura en lo que respecta a los parámetros antropométricos que son acústicamente relevantes en la inferencia de HRTF más precisas, además de la recopilación de nuevas y más extensas bases de datos de HRTF y antropometría.

En este trabajo se presenta un método para personalizar HRTF que se ahorra las desventajas prácticas y económicas de la medición directa sobre un sujeto dado. El método es modular, lo que permite la adición de mejoras futuras a medida en que se produzcan avances en el estado del arte en la personalización de HRTF,

grupo de parámetros antropométricos relevantes, y la creación de bases de datos más grandes y adecuadas para la tarea.

Un objetivo a largo plazo para el método presentado en este trabajo es su inclusión en dispositivos móviles, permitiéndoles a sus usuarios personalizar sus HRTF a partir de retratos obtenidos con el la cámara digital del mismo dispositivo.

Bibliografía

- [1] H. Moller, M.F. Sorensen, D. Hammershoi, and C.B. Jensen. Head related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43:300–321, May 1995.
- [2] J. Blauert. *Spatial Hearing*, pages 78–89. MIT Press, Cambridge, MA, 1997.
- [3] Corey I. Cheng and Gregory H. Wakefield. Introduction to Head-Related Transfer Functions (HRTF's): Representations of HRTF's in Time, Frequency, and Space. In *Audio Engineering Society Convention 107*, September 1999.
- [4] C.I. Cheng and G.H. Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. *J. Audio Eng. Soc.*, 49(4):231–249, 2001.
- [5] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3–4):171–218, 1992.
- [6] Ana Laura Padilla Ortíz, Felipe Orduña Bustamante, Rafael de la Guardia González, and José A. Córdova Magaña. Tecnologías de reproducción de sonido biaural a través de audífonos y su aplicación en dispositivos móviles de comunicación y entretenimiento. In *Congreso SOMI XXIII, Jalapa, México*, October 2008.
- [7] Colin E. Cherry. Some experiments on the recognition of speech, with one and two ears. *J. Acoust. Soc. Am.*, 25(5):975–979, 1953.
- [8] B. Nordlund and B. Fritzell. The influence of azimuth on speech signals. *Acta Oto-laryngologica*, 56:632–642, 1963.
- [9] J. P. Moncur and D. D. Dirks. Binaural and monaural speech intelligibility in reverberation. *J. Speech Hear. Res.*, 10:186–195, 1967.
- [10] A. L. Padilla-Ortiz and F. Orduña-Bustamante. Binaural speech intelligibility and interaural cross-correlation under disturbing noise and reverberation. *Journal of Applied Research and Technology*, 10(3):347–360, June 2012.
- [11] ANSI S3.36-1985 American National Standard Specification for a Manikin for Simulated In-Situ Airborne Acoustic Measurements, 1985.
- [12] W.G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97:3907–3908, 1995.
- [13] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman. Localization using non-individualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1):111–123, July 1993.
- [14] F. Wightman and D. Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878, February 1989.
- [15] V. R. Algazi, C. Avendano, and R. O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.*, 109(3):1110–1122, March 2001.
- [16] Ricardo R Boullosa and A Pérez López. Some acoustical properties of the anechoic chamber at the Centro de Instrumentos, Universidad Nacional

- Autónoma de México. *Applied Acoustics*, 56(3):199 – 207, 1999.
- [17] Dmitry N. Zotkin, Ramani Duraiswami, Elena Grassi, and Nail A. Gumerov. Fast head-related transfer function measurement via reciprocity. *J. Acoust. Soc. Am.*, 120(4):2202–2215, October 2006.
- [18] Bruno Sanches Masiero. *Individualized Binaural Technology. Measurement, Equalization and Perceptual Evaluation*. PhD thesis, Rheinischen–Westfälischen Technischen Hochschule, Aachen University, December 2012.
- [19] D. N. Zotkin, J. Hwang, R. Duraiswami, and L. S. Davis. HRTF personalization using anthropometric measurements. In *Proc. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2003)*, pages 157–160, October 2003.
- [20] V. R. Algazi, R. O. Duda, D. P. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. IEEE WAS PAA01*, pages 99–102, New Paltz, NY, October 2001.
- [21] D. Zotkin, R. Duraiswami, and L. Davis. Rendering localized spatial audio in a virtual auditory space. *IEEE Trans. Multimedia*, 6(4):553–564, August 2004.
- [22] T. Nishino, N. Inoue, K. Takeda, and F. Itakura. Estimation of HRTFs on the horizontal plane using physical features. *Applied Acoustics*, 68:897–908, 2007.
- [23] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.
- [24] Jonathon Shlens. A Tutorial on Principal Component Analysis. Technical report, Systems Neurobiology Laboratory, University of California at San Diego, 2005.
- [25] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno. Reconstructing head models from photographs for individualized 3d-audio processing. *Computer Graphics Forum*, 27(7):1719–1727, October 2008.
- [26] Ramani Duraiswami, Larry Davis, Shihab A. Shamma, Howard C. Elman, Richard O. Duda, V. Ralph Algazi, Qing-Ho Liu, and S. T. Raveendra. Individualized HRTFs using computer vision and computational acoustics. *Journal of The Acoustical Society of America*, 108, 2000.
- [27] Pablo Lapunzina and Horacio Aiello. *Manual de antropometría normal y patológica. Fetal, neonatal, niños y adultos*. Masson, Barcelona, España., 2002.
- [28] Ali Abbas. The role of the human ear in the process of forensic identification. Master’s thesis, Leicester Royal Infirmary, 2003.
- [29] M. D. Burkhard and R.M. Sachs. Anthropomorphic manikin for acoustic research. *J. Acoust. Soc. Am.*, 58:214–222, 1975.
- [30] U. Burandt, C. Posselt, S. Ambrozis, M. Hosenfeld, and V. Knauff. Anthropometric contribution to standardising mannequins for artificial-head microphones and to measuring headphones and ear protectors. *Applied Ergonomics*, pages 373–378, 1991.
- [31] Jkrome E. Adamson, Charles E. Hortox, and Hugh H. Crawford. The growth pattern of the external ear. *Plastic and Reconstructive Surgery*, 36:466–470, 1965.
- [32] Milton Alexander and Lloyd L. Laubach. Anthropometry of the human ear (a photogrammetric study of usaf flight personnel). Technical report, Aerospace Medical Research Laboratories, January 1968.
-

- [33] L. G. Farkas. Vertical location of the ear, assessed by the leiber test, in healthy north american caucasians 6–19 years of age. *Arch Otorhinolaryngol.*, 220(1–2):9–13, March 1978.
- [34] L. G. Farkas. Anthropometry of normal and anomalous ears. *Clinics in Plastic Surgery*, 5:401–412, 1978.
- [35] Eugene H. Courtiss and L. G Farkas. Anthropometry of normal and anomalous ears. *Plastic and Reconstructive Surgery*, 63, 1979.
- [36] R. Azaria, N. Adler, R. Silfen, D. Regev, and D. J. Hauben. Morphometry of the adult human earlobe: A study of 547 subjects and clinical application. *Plastic and Reconstructive Surgery*, 111(7):2398–2402, June 2003.
- [37] C. Barut and E. Aktunc. Anthropometric measurements of the external ear in a group of turkish primary school students. *Aesthetic Plastic Surgery*, 30:255–259, 2006.
- [38] Bernd Jähne, Horst Haußecker, and Peter Geißler. *Handbook of Computer Vision and Applications*, volume 2, Signal Processing and Pattern Recognition. Academic Press, 1999.
- [39] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2nd. edition, 2002.
- [40] Erik Hjeltnäs and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83:236–274, 2001.
- [41] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, May 1995.
- [42] I. J. Cox, J. Ghosn, and P. N. Yianilos. Feature-based face recognition using mixture-distance. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, 1996.
- [43] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.
- [44] Kevin W. Bowyer, Kyong Chang, and Patrick Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.
- [45] Hui Chen and B. Bhanu. Human ear detection from side face range images. In *Proceedings of the 17th International Conference on Pattern Recognition. ICPR 2004.*, volume 3, pages 574–577, August 2004.
- [46] B. Moreno, A. Sanchez, and J. Velez. On the use of outer ear images for personal identification in security applications. In *IEEE International Carnahan Conference on Security Technology*, pages 469–476, 1999.
- [47] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama. Study on individual recognition for ear images by using genetic local search. In *Proceedings of the 2002 Congress on Evolutionary Computation*, pages 237–242, 2002.
- [48] Bir Bhanu and Hui Chen. Human ear recognition in 3d. In *Workshop on Multimodal User Authentication*, pages 91–98, 2003.
- [49] Ping Yan and Kevin W. Bowyer. ICP-based approaches for 3D ear recognition. In *Biometric Technology for Human Identification II, Proceedings of SPIE*, volume 5779, pages 282–291, 2005.
- [50] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *IJIVC*, 10(3):145–155, 1992.
-

-
- [51] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2):239–256, 1992.
- [52] Bir Chen, Hui and Bhanu. Contour matching for 3d ear recognition. In *Seventh IEEE Workshops on Application of Computer Vision. WACV/MOTIONS '05.*, volume 1, pages 123–128, January 2005.
- [53] D. Hurley, M. Nixon, and J. Carter. Force field energy functionals for image feature extraction. *Image and Vision Computing Journal*, 20:429–432, 2002.
- [54] David J. Hurley, Mark S. Nixon, and John N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98(3):491–512, June 2005.
- [55] Sepehr Attarchi, Karim Faez, and Aref Rafiei. A new segmentation approach for ear recognition. In Jacques Blanc-Talon, Salah Bourennane, Wilfried Phillips, Dan Popescu, and Paul Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 5259 of *Lecture Notes in Computer Science*, pages 1030–1037. Springer Berlin / Heidelberg, 2008. 10.1007/978-3-540-88458-3_93.
- [56] Sepehr Attarchi, Masoud Nosrati, and Karim Faez. A new ear recognition approach for personal identification. In De-Shuang Huang, Donald Wunsch, Daniel Levine, and Kang-Hyun Jo, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, volume 5226 of *Lecture Notes in Computer Science*, pages 356–363. Springer Berlin / Heidelberg, 2008. 10.1007/978-3-540-87442-3_45.
- [57] Bir Bhanu and Hui Chen. Ear detection and recognition in 2d and 3d. In *Human Ear Recognition by Computer*, Advances in Pattern Recognition, pages 11–19. Springer London, 2008. 10.1007/978-1-84800-129-9_2.
- [58] Hui Chen and B. Bhanu. Shape model-based 3d ear detection from side face range images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 122, 2005.
- [59] Hui Chen and Bir Bhanu. Human ear recognition in 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):718–737, April 2007.
- [60] H. Chen and B. Bhanu. Human ear detection from 3d side face range images. In Andreas Koschan, Marc Pollefeys, and Mongi Abidi, editors, *3D Imaging for Safety and Security*, volume 35 of *Computational Imaging and Vision*, pages 133–155. Springer Netherlands, 2007.
- [61] Mohamed Abdel-Mottaleb and Jindan Zhou. Human ear recognition from face profile images. In David Zhang and Anil Jain, editors, *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 786–792. Springer Berlin / Heidelberg, 2005.
- [62] T. J. Coward, B. J. J. Scott, R. M. Watson, and R. Richards. Identifying the position of an ear from a laser scan: The significance for planning rehabilitation. *Int. J. Oral Maxillofac. Surg.*, 31:244–251, 2002.
- [63] Ping Yan and Kevin W. Bowyer. An automatic 3d ear recognition system. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 326–333. IEEE Computer Society, June 2006.
- [64] Zouhour Ben Azouz, Chang Shu, and Anja Mantel. Automatic locating of anthropometric landmarks on 3d human
-

- models. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, 3DPVT '06, pages 750–757, Washington, DC, USA, 2006. IEEE Computer Society.
- [65] Reyes Enciso, Alex Shaw, Ulrich Neumann, and James Mah. 3D head anthropometric analysis. In *proceedings of the international society for optical...*, 2003.
- [66] J.M. Flores, I. Bloch, T. Bousquet, F. Schmitt, and C. Grangeat. Shape-based averaging for craniofacial anthropometry. In *Computer Science, 2005. ENC 2005. Sixth Mexican International Conference on*, pages 314–319, 2005.
- [67] M. Burge and W. Burger. Ear biometrics for machine vision. In *21th Workshop of the Austrian Association for Pattern...*, 1997.
- [68] Mark Burge and Wilhelm Burger. Ear biometrics. In Anil K. Jain, Ruud Bole, and Sharath Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, The International Series in Engineering and Computer Science. Springer, January 1999.
- [69] M. Burge and W. Burger. Ear biometrics in computer vision. In *Proceedings. 15th International Conference on Pattern Recognition.*, volume 2, pages 822–826, 2000.
- [70] Michał Choras. Ear biometrics based on geometrical feature extraction. *Electronic Letters on Computer Vision and Image Analysis*, 5(3):84–95, 2005.
- [71] Kyong Chang, K.W. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1160–1165, September 2003.
- [72] Ankur Mohan, Ramani Duraiswami, Dmitry N. Zotkin, Daniel DeMenthon, and Larry S. Davis. Using computer vision to generate customized spatial audio. In *ICME 2003*, 2003.
- [73] Jens Blauert. *Spatial hearing. The psychophysics of human sound localization*. The MIT Press, Cambridge, MA, 1997.
- [74] Lord Rayleigh (J. W. Strutt. 3rd Baron of Rayleigh). On our perception of sound direction. *Philos. Mag.*, 13:214–232, 1907.
- [75] G. W. Stewart. The function of intensity and phase in the binaural location of pure tones. *Phys. Rev.*, page 425, May 1920.
- [76] Hartley and Fry. The binaural location of pure tones. *Phys. Rev.*, page 440, December 1921.
- [77] F. Firestone. The phase difference and amplitude ratio at the ears due to a source of pure tone. *J. Acoust. Soc. Am.*, 2:260–270, 1930.
- [78] V. R. Algazi, R. O. Duda, and D. M. Thompson. The use of head-and-torso models for improved spatial sound synthesis. In *Proc. AES 113th Convention, Los Angeles, CA*, 2002.
- [79] A. Kulkarni, S. K. Isabelle, and H. S. Colburn. Sensitivity of human subjects to head-related transfer-function phase spectra. *J. Acoust. Soc. Am.*, 105(5):2821–2840, 1999.
- [80] Vikas C. Raykar, Ramani Duraiswami, Larry Davis, and B. Yegnanarayana. Extracting significant features from the HRTF. In *Proc. of the 2003 International Conference on Auditory Display, Boston, MA, USA*, July 2003.
- [81] Vikas C. Raykar, Ramani Duraiswami, and B. Yegnanarayana. Extracting frequencies of the pinna spectral notches in
-

- measured head related impulse responses. 2004.
- [82] Vikas C. Raykar, Ramani Duraiswami, and B. Yegzanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *J. Acoust. Soc. Am.*, 118(1):364–374, July 2005.
- [83] MAkoto Otani and Tatsuya Hirahara. Numerical study on source distance dependency of head-related transfer functions. In *ICA 2007*, September 2007.
- [84] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Hearing distance: A low-cost model for near-field binaural effects. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 2030–2034. IEEE, 2012.
- [85] Paul M Hofman, Jos GA Van Riswick, and A John Van Opstal. Relearning sound localization with new ears. *Nature neuroscience*, 1(5):417–421, 1998.
- [86] M. P. Zwiers, A. J. Van Opstal, and J. R. M. Cruysberg. A spatial hearing deficit in early-blind humans. *The Journal of Neuroscience*, 21(RC142):1–5, 2001.
- [87] M Zwiers, Al Van Opstal, and J Cruysberg. Two-dimensional sound-localization behavior of early-blind humans. *Experimental Brain Research*, 140(2):206–222, 2001.
- [88] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Estimation and modeling of pinna-related transfer functions. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010.
- [89] Navarun Gupta and Armando Barreto. A Customizable Model of Head-Related Transfer Functions Based on Pinna Measurements. *International Journal of Scientific and Statistical Computing (IJSSC)*, 1(2), 2011.
- [90] E. A. López-Poveda and R. Meddis. A Physical Model of Sound Diffraction and Reflections in the Human Concha. *J. Acoust. Soc. Am.*, 100:3248–3259, 1996.
- [91] Sergio G. Rodríguez and Miguel A. Ramírez. Extracting and modeling approximated pinna-related transfer functions from hrtf data. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland*, July 2005.
- [92] Patrick Satarzadeh, V. Ralph Algazi, and Richard O. Duda. Physical and filter pinna models based on anthropometry. In *AES 122nd Convention, Vienna, Austria*, May 2007.
- [93] S. Spagnol, M. Geronazzo, and F. Avanzini. Fitting pinna-related transfer functions to anthropometry for binaural sound rendering. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 194–199, oct. 2010.
- [94] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Structural modeling of pinna-related transfer functions. 2010.
- [95] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. A head-related transfer function model for real-time customized 3-d sound rendering. *Signal-Image Technologies and Internet-Based System, International IEEE Conference on*, 0:174–179, 2011.
- [96] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Customized 3d sound for innovative interaction design. In *Proc. Italian ACM SigCHI Conf. on Computer-Human Interaction (CHI-taly’2011)*, Alghero, 2011.
- [97] S. Spagnol, M. Geronazzo, and F. Avanzini. On the relation between pinna reflection patterns and head-related transfer function features. *Audio, Speech*,

- and Language Processing, IEEE Transactions on*, 21(3):508–519, 2013.
- [98] J. Blauert. The AUDIS catalog of human HRTFs. In *Proc. 16th ICA, Seattle, WA*, pages 2901–2902, June 1998.
- [99] Shouichi Takane, Daisuke Arai, Tohru Miyajima, Kanji Watanabe, Yôiti Suzuki, and Toshio Sone. A database of head-related transfer functions in whole directions on upper hemisphere. *Acoust. Sci. and Tech.*, 23(3):160–162, 2002.
- [100] Olivier Warusfel. IRCAM HRTF Database.
- [101] BoSun Xie, XiaoLi Zhong, Dan Rao, and ZhiQiang Liang. Head-related transfer function database and its analyses. *Science in China Series G: Physics, Mechanics and Astronomy*, 50(3):267–280, June 2007.
- [102] Navarun Gupta, Armando Barreto, Manan Joshi, and Juan Carlos Agudelo. HRTF database at FIU DSP lab. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 169–172. IEEE, 2010.
- [103] Simone Spagnol, Marko Hiipakka, and Ville Pulkki. A single-azimuth pinna-related transfer function database. In *14th Int. Conf. Digital Audio Effects (DAFx-11)*, 2011.
- [104] Hagen Wierstorf, Matthias Geier, and Sascha Spors. A free database of head related impulse response measurements in the horizontal plane with multiple distances (Poster). In *Audio Engineering Society Convention 130*, 2011.
- [105] Austrian Academy of Sciences Acoustics Research Institute. ARI HRTF Database, 2013.
- [106] Areti Andreopoulou, Agnieszka Rogińska, and Hariharan Mohanraj. A database of repeated head-related transfer function measurements. In *International Conference on Auditory Display (ICAD)*, 2013.
- [107] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867, February 1989.
- [108] K. A. J. Riederer. Head-related transfer function measurement. Master’s thesis, Helsinki University of Technology, Helsinki, 1998.
- [109] B. P. Bovbjerg, F. Christensen, and P. Minnaar. Measuring the head related transfer functions of an artificial head with a high directional resolution. In *AES 109th Convention*, Los Angeles, California, 2000.
- [110] P Guillon, R Zolfaghari, N Epain, A van Schaik, CT Jin, C Hetherington, J Thorpe, and A Tew. Creating the sydney york morphological and acoustic recordings of ears database. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 461–466. IEEE, 2012.
- [111] Agnieszka Andreopoulou, Areti;Roginska. Towards the creation of a standardized hrtf repository. In *Audio Engineering Society Convention 131*, 10 2011.
- [112] Michele Geronazzo, Fabrizio Granza, Simone Spagnol, and Federico Avanzini. A standardized repository of head-related and headphone impulse response data. In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.
- [113] Klaus AJ Riederer. Repeatability analysis of head-related transfer function measurements. In *Audio Engineering Society*
-

- Convention 105*. Audio Engineering Society, 1998.
- [114] Marko Hiipakka, Teemu Kinnari, and Ville Pulkki. Estimating head-related transfer functions of human subjects from pressure-velocity measurements. *The Journal of the Acoustical Society of America*, 131:4051, 2012.
- [115] LJP Van der Maaten. An introduction to dimensionality reduction using matlab. Technical Report MICC 07-07, Faculty of Humanities & Sciences, MICC/IKAT, Maastricht, The Netherlands, July 2007.
- [116] LJP Van der Maaten, EO Postma, and HJ Van Den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1-41, 2009.
- [117] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91:1637-1647, 1992.
- [118] J. Middlebrooks and D. Green. Observations on a principal components analysis of head related transfer functions. *J. Acoust. Soc. Am.*, 92(1):597-599, 1992.
- [119] Lin Wang, Fuliang Yin, and Zhe Chen. A hybrid compression method for head-related transfer functions. *Applied Acoustics*, 70(9):1212-1218, September 2009.
- [120] J.C.B. Torres, M.R. Petraglia, and R.A. Tenenbaum. HRTF modeling for efficient auralization. In *IEEE International Symposium on Industrial Electronics, 2003. ISIE '03.*, volume 2, pages 919-923, June 2003.
- [121] Julio Cesar B. Torres, Mariane R. Petraglia, and Roberto A. Tenenbaum. An efficient wavelet-based HRTF model for auralization. *Acta Acustica united with Acustica*, 90(1):108-120, January/-February 2004.
- [122] J.C.B. Torres, M.R. Petraglia, and R.A. Tenenbaum. Low-order modeling and grouping of hrtfs for auralization using wavelet transforms. In *IEEE International Conference on Acoustics, Speech, and Signal Processing. (ICASSP '04).*, volume 4, May 2004.
- [123] J.C.B. Torres, M.R. Petraglia, and A. Petraglia. Efficient structure for auralization using a wavelet-based HRTF model. In J. Silva-Martinez, editor, *Circuits, Signals, and Systems (CSS 2006)*, San Francisco, CA, USA, 2006.
- [124] Julio Cesar B Torres, Roberto A Tenenbaum, and Mariane R Petraglia. Efficient auralization by grouping directions and modeling HRTF's using wavelets. In *Proc. XII International Symposium on Dynamic Problems of Mechanics, Ilhabela*, 2007.
- [125] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323-2326, December 2000.
- [126] Ramani Duraiswami and Vikas C. Raykar. The manifolds of spatial hearing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP05*, pages 285-288, 2005.
- [127] Vikas C. Raykar and Ramani Duraiswami. On the manifolds of spatial hearing. In *NIPS 2006 workshop on Novel Applications of Dimensionality Reduction*, 2006.
- [128] KJ Faller, Armando Barreto, and Malek Adjouadi. Decomposition of head-related transfer functions based on the hankel total least squares method. In *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th*, pages 161-166. IEEE, 2009.
-

-
- [129] Kenneth John Faller II, Armando Barreto, and Malek Adjouadi. Augmented Hankel Total Least-Squares Decomposition of Head-Related Transfer Functions. *J. Audio Eng. Soc.*, 58(1/2):3–21, January 2010.
- [130] Claire Masterson, Stephen Adams, Gavin Kearney, and Frank Boland. A method for head related impulse response simplification. In *17th European Signal Processing Conference (EUSIPCO 2009)*, pages 2569–2573, Glasgow, Scotland, August 2009.
- [131] Martin Rothbucher, Hao Shen, and Klaus Diepold. Dimensionality reduction in hrtf by using multiway array analysis. In Helge Ritter, Gerhard Sagerer, Rüdiger Dillmann, and Martin Buss, editors, *Human Centered Robot Systems*, volume 6 of *Cognitive Systems Monographs*, pages 103–110. Springer Berlin Heidelberg, 2009.
- [132] Wen Zhang, Thushara D. Abhayapala, Rodney A. Kennedy, and Ramani Duraiswami. Insights into head-related transfer function: Spatial dimensionality and continuous representation. *J. Acoust. Soc. Am.*, 127(4):2347–2357, April 2010.
- [133] Yongqing Tang, Yong Fang, and Qinghua Huang. Audio personalization using head related transfer function in 3dtv. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, pages 1–4, 2011.
- [134] Areti Andreopoulou, Agnieszka Roginska, and Juan Pablo Bello. Reduced representations of hrtf datasets: A discriminant analysis approach. In *Audio Engineering Society Convention 135*. Audio Engineering Society, 2013.
- [135] Jeroen Breebaart, Fabian Nater, and Armin Kohlrausch. Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing. *J. Audio Eng. Soc.*, 58(3):126–140, March 2010.
- [136] S Carlile, C Jin, and V Van Raad. Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia*, pages 220–223, 2000.
- [137] Ze-Wei Chen, Guang-Zheng Yu, Bo-Sun Xie, and Shan-Qun Guan. Calculation and analysis of near-field head-related transfer functions from a simplified head-neck-torso model. *Chinese Physics Letters*, 29(3):034302, 2012.
- [138] Corey I. Cheng. *Visualization, Measurement, and Interpolation of Head-Related Transfer Functions (HRTF's) with Applications in Electro-Acoustic Music*. PhD thesis, The University of Michigan, 2001.
- [139] Ramani Duraiswami, Dmitry N. Zotkin, and Nail A. Gumerov. Interpolation and range extrapolation of hrtfs. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, May 2004*.
- [140] Ramani Duraiswami, Dmitry N. Zotkin, Zhiyun Li, Elena Grassi, Nail A. Gumerov, and Larry S. Davis. High order spatial audio capture and its binaural head-tracked playback over headphones with hrtf cues. In *Proc. AES 119th Convention, New York, NT, USA, October 2005*.
- [141] Simone Fontana, Angelo Farina, and Yves Grenier. A system for head related impulse responses rapid measurement and direct customization. In *AES 120th Convention, Paris, France, May 2006*.
- [142] Hannes Gamper. Selection and interpolation of head-related transfer functions
-

- for rendering moving virtual sound sources. In *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, September 2–5 2013.
- [143] Hannes Gamper. Head-related transfer function interpolation in azimuth, elevation, and distance. *The Journal of the Acoustical Society of America*, 134(6):EL547–EL553, 2013.
- [144] Klaus Hartung, Jonas Braasch, and Susanne J Sterbing. Comparison of different methods for the interpolation of head-related transfer functions. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, March 1999.
- [145] Yang Liu and Bosun Xie. Analysis on the stability of spatial interpolation schemes for head-related transfer function. *The Journal of the Acoustical Society of America*, 131(4):3305–3305, 2012.
- [146] Yuancheng Luo, Dmitry N Zotkin, Hal Daume, and Ramani Duraiswami. Kernel regression for head-related transfer function interpolation and spectral extrema extraction. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 256–260. IEEE, 2013.
- [147] Takanori Nishino, Sumie Mase, Shoji Kajita, Kazuya Takeda, and Fumitada Itakura. Interpolating HRTF for auditory virtual reality. *J. Acoust. Soc. Am.*, 100(4):2602–2602, October 1996.
- [148] T. Nishino, S. Kajita, K. Takeda, and F. Itakura. Interpolation of the head related transfer function on the horizontal plane. *J. Acoust. Soc. Jpn. (in Japanese)*, 55(2):91–99, 1999.
- [149] Takanori Nishino, Mikio Ikeda, Kazuya Takeda, and Fumitada Itakura. Interpolating head related transfer functions. In *The Seventh Western Pacific Regional Acoustics Conference. WESTPRA VII*, Kumamoto, Japan, 3-5 October, 2000.
- [150] German Ramos and Maximo Cobos. Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications. *The Journal of the Acoustical Society of America*, 134(3):1735–1738, 2013.
- [151] S. Spors and J. Ahrens. Efficient range extrapolation of head-related impulse responses by wave field synthesis techniques. In *IEEE International Conference on Acoustics, Speech, and Signal Processing. (ICASSP '11)*, Prague, Czech Republic, 2011.
- [152] Kanji Watanabe, Shouichi Takane, and Yooiti Suzuki. Interpolation of head-related transfer functions based on the common-acoustical-pole and residue model. *Acoustical Science and Technology*, 24(5):335–337, 2003.
- [153] Kanji Watanabe, Shouichi Takane, and Yôiti Suzuki. A novel interpolation method of hrtfs based on the common-acoustical-pole and zero model. *Acta Acustica united with Acustica*, 91(6):958–966, November/December 2005.
- [154] Xiaoli Zhong. Interpolation of head-related transfer functions using neural network. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on*, volume 2, pages 565–568. IEEE, 2013.
- [155] Zoltan Haraszy, Daniel Ianchis, Virgil Tiponut, NE Mastorakis, V Mladenov, Z Bojkovic, S Kartalopoulos, A Varonides, and M Jha. Generation of the head related transfer functions using artificial neural networks. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, number 13. World Scientific and Engineering Academy and Society, 2009.
-

-
- [156] Zoltan Haraszy, Sebastian Micut, Virgil Tiponut, and Titus Slavici. Multi-subject head related transfer function generation using artificial neural networks. *Measurement*, 1, 2010.
- [157] Zoltan Haraszy, David-George Cristea, Virgil Tiponut, and Titus Slavici. Improved head related transfer function generation and testing for acoustic virtual reality development. In *7th WSEAS International Conference on Engineering Education*, pages 22–24, 2010.
- [158] Zoltan Haraszy, David-George Cristea, Sebastian Micut, and Ivan Bogdanov. Efficient algorithm for extracting essential head related impulse response data for acoustic virtual reality development. In *Proceedings of the 15th WSEAS international conference on Systems*, pages 315–320. World Scientific and Engineering Academy and Society (WSEAS), 2011.
- [159] Y. Iwaya. Individualization of head-related transfer functions with tournament-style listening test: Listening with other’s ears. *Acoustical Science and Technology*, 27(6):340–343, 2006.
- [160] Sungmok Hwang and Youngjin Park. HRIR customization in the median plane via principal components analysis. In *AES 31st International Conference*, pages 1–9, London, UK, June 2007.
- [161] Josef Hölzl. An initial investigation into HRTF Adaptation using PCA. Master’s thesis, Institut für Elektronische Musik und Akustik, July 2012.
- [162] Richard H. Y. So, N. M. Leung, Andrew B. Horner, Jonas Braasch, and K. L. Leung. Effects of spectral manipulation on nonindividualized head-related transfer functions (HRTFs). *Human Factors: The Journal of the Human Factors and Ergonomics Society*, June 2011.
- [163] Rozenn Nicol, Vincent Lemaire, Alexis Bondu, and Sylvain Busson. Looking for a relevant similarity criterion for HRTF clustering: a comparative study. In *Proc. AES 120th Convention, Paris, France*, May 2006.
- [164] Song Xu, Zhizhong Li, and Gavriel Salvendy. Individualized head-related transfer functions based on population grouping (1). *J. Acoust. Soc. Am.*, 124(5):2708–2710, November 2008.
- [165] R. H.Y. So, B. Ngan, A. Horner, J. Braasch, J. Blauert, and K. L. Leung. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. *Ergonomics*, 53(6):767–781, May 2010.
- [166] Areti Andreopoulou, Agnieszka Roginska, and Juan Bello. Observing the clustering tendencies of head related transfer function databases. In *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.
- [167] J. C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, 106(3):1480–1492, September 1999.
- [168] J. C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.*, 106(3):1493–1510, 1999.
- [169] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan. Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Am.*, 108(6):3088–3091, 2000.
- [170] V. Ralph Algazi, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang. Approximating the head-related transfer function using simple
-

- geometric models of the head and torso. *J. Acoust. Soc. Am.*, 112(5):2053–2064, 2002.
- [171] Carlos Avendano, V. Ralph Algazi, and Richard O. Duda. A head-and-torso model for low-frequency binaural elevation effects. In *Proc. 1999 IEEE Workshop on Applications of Signals Processing to Audio and Acoustics*, pages 179–182, New Paltz, NY, October 1999.
- [172] Carlos Avendano, Richard O. Duda, and V. Ralph Algazi. Modeling the contralateral HRTF. In *AES 16th International Conference on Spatial Sound Reproduction*, 1999.
- [173] Richard O. Duda and William L. Martens. Range-dependence of the hrtf for a spherical head. 1997.
- [174] Richard O. Duda and William L. Martens. Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.*, 104(5):3048–3058, November 1998.
- [175] Richard O. Duda, Carlos Avendano, and V. Ralph Algazi. An adaptable ellipsoidal head model for the interaural time difference. In *IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP 99*, 1999.
- [176] Richard S. Juszkievicz. An improvement to anthropometry-based head and torso hrtf synthesis models for locations near the frontal median plane. Master’s thesis, University of Miami, 2007.
- [177] Guang-Zhen Yu, Bo-Sun Xie, and Rao Dan. Effect of sound source scattering on measurement of near-field head-related transfer functions. *Chinese Physics Letters*, 25(8):2926–2929, 2008.
- [178] Harald Ziegelwanger and Piotr Majdak. Modeling the direction-continuous time-of-arrival in head-related transfer functions. *The Journal of the Acoustical Society of America*, 135(3):1278–1293, 2014.
- [179] C. Phillip Brown and Richard O. Duda. An efficient HRTF model for 3-D sound. 1997.
- [180] C. Phillip Brown and Richard O. Duda. A structural model for binaural sound synthesis. *IEEE Transactions on Speech and Audio Processing*, 6(5):476–488, September 1998.
- [181] Richard O. Duda. Modeling head related transfer functions. In *Twenty-Seventh Asilomar Conference on Signals, Systems and Computers.*, Asilomar, CA., October–November 1993.
- [182] Jyri Huopaniemi and Matti Karjalainen. Hrtf filter design based on auditory criteria. In *Nordic Acoustical Meeting*, pages 323–330, Helsinki, June 1996.
- [183] Norman Herkamp Adams. *A Model of Head-Related Transfer Functions based on a State-Space Analysis*. PhD thesis, The University of Michigan, 2008.
- [184] D.N. Zotkin, R. Duraiswami, and N.A. Gumerov. Regularized hrtf fitting using spherical harmonics. In *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on*, pages 257–260, 2009.
- [185] Navarun Gupta, Armando Barreto, and Maroof Choudhury. Modeling head-related transfer functions based on pinna anthropometry. In *LACCEI'2004*, 2004.
- [186] Kanji Watanabe, Kenji Ozawa, Yukio Iwaya, Yôiti Suzuki, and Kenji Aso. Estimation of interaural level difference based on anthropometry and its effect on sound localization. *J. Acoust. Soc. Am.*, 122(5):2832–2841, November 2007.
- [187] Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi. Boundary element simulation of HRTFs and sound fields produced by virtual acoustic imaging systems. In *Proceedings of the 105th AES Convention, San Francisco*, 1998.
-

-
- [188] Yuvi Kahana, Philip A. Nelson, Maurice Petyt, and Sunghoon Choi. Numerical modelling of the transfer functions of a dummy-head and of the external ear. In *16th International Conference: Spatial Sound Reproduction*, March 1999.
- [189] Yuvi Kahana. *Numerical modelling of the head-related transfer function*. PhD thesis, University of Southampton, 2000.
- [190] Yuvi Kahana and Philip A. Nelson. Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. *Journal of sound and vibration*, 300:552–579, 2007.
- [191] Brian F. G. Katz. Boundary element method calculation of individual head-related transfer function. I. rigid model calculation. *J. Acoust. Soc. Am.*, 110(5):2440–2448, 2001.
- [192] Brian F. G. Katz. Boundary element method calculation of individual head-related transfer function. II. impedance effects and comparisons to real measurements. *J. Acoust. Soc. Am.*, 110(5):2449–2455, 2001.
- [193] K. Terai and I. Kakuhari. HRTF calculation with less influence from 3-D modeling error: Making a physical human head model from geometric 3-d data. *Acoustical Science and Technology*, 24(5):333–334, 2003.
- [194] T. Walsh, L. Demkowicz, and Charles R. Boundary Element Modeling of External Human Auditory System. *J. Acoust. Soc. Am.*, 115:1033–1043, 2004.
- [195] H Ziegelwanger, P Majdak, and A Reichinger. Calculation of listener-specific head-related transfer functions: Effect of mesh quality. *The Journal of the Acoustical Society of America*, 133(5):3282–3282, 2013.
- [196] Tian Xiao and Qing Huo Liu. Finite difference computation of head-related transfer function for human hearing. *J. Acoust. Soc. Am.*, 113(5):2434–2441, May 2003.
- [197] Mingsian R. Bai and Teng-Chieh Tsao. Numerical modeling of head-related transfer functions using the boundary source representation. *Journal of vibration and acoustics*, 128(5):594–603, 2006.
- [198] D. N. Zotkin, R. Duraiswami, L. S. Davis, A. Mohan, and V. Raykar. Virtual audio system customization using visual matching of ear parameters. In *Proc. IEEE ICPR 2002, Quebec City, Canada*, volume 3, pages 1003–1006, 2002.
- [199] V. Ralph Algazi, Pierre L. Divenyi, V. A. Martinez, and Richard O. Duda. Subject dependent transfer functions in spatial hearing. In *IEEE MWSCAS*, 1997.
- [200] Mingsian R. Bai and Kwuen-Yieng Ou. Head-related transfer function (HRTF) synthesis based on a three-dimensional array model and singular value decomposition. *Journal of Sound and Vibration*, 281:1093–1115, March 2005.
- [201] Oscar A. Ramos, Goncal Calvo, and Fabián C. Tommasini. Modelo acústico de cabeza y torso mediante análisis de componentes principales. In *Mecánica Computacional Vol XXVI*, pages 46–58, October 2007.
- [202] T Nishino, Y Nakai, K Takeda, and F. Itakura. Estimating head related transfer function using multiple regression analysis. In *IEICE Trans A*, volume J84-A(3), pages 260–8, 2001.
- [203] N. Inoue, T. Kimura, T. Nishino, K. Itou, and K. Takeda. Evaluation of HRTFs estimated using physical features. *Acoustical Science and Technology*, 26(5):453–455, 2005.
-

- [204] Sergio G. Rodríguez and Miguel A. Ramírez. Linear relationships between spectral characteristics and anthropometry of the external ear. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, 2005*, pages 336–339, July 2005.
- [205] Hongmei Hu, Lin Zhou, Jie Zhang, Hao Ma, and Zhenyang Wu. Head related transfer function personalization based on multiple regression analysis. In *Computational Intelligence and Security, 2006 International Conference on*, November 2006.
- [206] Graham Grindlay and M Alex O Vasilescu. A multilinear approach to hrtf personalization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP 07*, 2007.
- [207] Hong-Mei Hu, Lin Zhou, Hao Ma, and Zhen-Yang Wu. Head-related transfer function personalization based on partial least square regression. *Journal of Electronics and Information Technology*, 30(1):154–158, January 2008.
- [208] Song Xu, Zhizhong Li, and Gavriel Salvendy. Improved method to individualize head-related transfer function using anthropometric measurements. *Acoustical Science and Technology*, 29(6):388–390, 2008.
- [209] Hugeng, Wahidin Wahab, and Dadang Gunawan. Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements. *Journal of Telecommunications*, 2(2):31–41, may 2010.
- [210] Hugeng, Wahab W., and D. Gunawan. Enhanced individualization of head-related impulse response model in horizontal plane based on multiple regression analysis. In *Computer Engineering and Applications (ICCEA), 2010 Second International Conference on*, pages 226 – 230, March 2010.
- [211] Martin Rothbucher, Marko Durkovic, Hao Shen, and Klaus Diepold. HRTF customization using multiway array analysis. In *18th European Signal Processing Conference (EUSIPCO-2010)*, pages 229–233, Aalborg, Denmark, August 2010.
- [212] Xiang-Yang Zeng, Shu-Guang Wang, and Li-Ping Gao. A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures. *Journal of sound and vibration*, 329(19):4093–4106, September 2010.
- [213] Hongmei Hu, Lin Zhou, Hao Ma, and Zhenyang Wu. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, 69:163–172, 2008.
- [214] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile. Enabling individualized virtual auditory space using morphological measurements. In *Proceedings of the First IEEE Pacific-Rim Conf. on Multimedia*, pages 235–238, 2000.
- [215] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson. Structural composition and decomposition of HRTFs. In *Proc. IEEE WASPAA01, New Paltz, NY*, pages 103–106, 2001.
- [216] Janina Fels and Michael Vorländer. Anthropometric parameters influencing head-related transfer functions. *Acta Acustica united with Acustica*, 95(2):331–342, March/April 2009.
- [217] Hugeng, Wahidin Wahab, and Dadang Gunawan. The effectiveness of chosen partial anthropometric measurements in individualizing head-related transfer functions on median plane. *ITB Journal of Information and Communication Technology*, 5(1):35–56, 2011.
-

- [218] M. Zhang, R. A. Kennedy, T. D. Abhayapala, and W. Zhang. Statistical method to identify key anthropometric parameters in HRTF individualization. In *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, 2011.
- [219] Simone Spagnol, Davide Rocchesso, Michele Geronazzo, and Federico Avanzini. Automatic extraction of pinna edges for binaural audio customization. In *Multi-media Signal Processing (MMSP), 2013 IEEE 15th International Workshop on*, pages 301–306. IEEE, 2013.
- [220] Simone Spagnol, Michele Geronazzo, Davide Rocchesso, and Federico Avanzini. Extraction of pinna features for customized binaural audio delivery on mobile devices. In *Proceedings of MoMM2013*, 2013.
- [221] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.
- [222] P. Mahalanobis. On tests and measures of group divergence I. Theoretical formulae. *J. and Proc. Asiat. Soc. of Bengal*, 26:541—588, 1930.
- [223] Ghassan Hamarneh, Rafeef Abu-Gharbieh, and Tomas Gustavsson. Active shape models - part i: Modeling shape and gray level variations. In *Proceedings of the Swedish Symposium on Image Analysis, SSAB*, 1998.
- [224] Rafeef Abu-Gharbieh, Ghassan Hamarneh, and Tomas Gustavsson. Active shape models - part ii: Image search and classification. In *Proceedings of the Swedish Symposium on Image Analysis, SSAB*, 1998.
- [225] T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. Technical report, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester, 2001.
- [226] F. Arámbula Cosío, J.A. Márquez Flores, and M.A. Padilla Castañeda. Use of simplex search in active shape models for improved boundary segmentation. *Pattern Recognition Letters*, 31:806—817, 2010.
- [227] Eugene Hecht. *Óptica*. Addison Wesley, Madrid, 3rd edition, 2000.
- [228] M. Sonka and J.M. Fitzpatrick, editors. *Handbook of Medical Imaging*, volume 2. SPIE Press, 2000.
- [229] ANSI. ANSI/ASA S1.4-2014, Electroacoustics - Sound Level Meters - Part 1: Specifications. American National Standards Institute, New York, 2014.
- [230] IEC. IEC 61672-1:2013 Ed. 2.0 Electroacoustics - Sound level meters, 2013.
- [231] Hugeng, Wahidin Wahab, and Dadang Gunawan. Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis. *Signal Processing: An International Journal (SPIJ)*, 4(4):201–212, 2010.
- [232] Martin Hagan, Howard Demuth, and Mark Beale. *Neural Network Design*. Brooks/Cole Publishing Company, 1996.
- [233] Howard Demuth and Mark Beale. *Neural Network Toolbox*. The MathWorks, 4 edition, 2002.
- [234] Mark Beale, Martin T. Hagan, and Howard Demuth. *Neural Network Toolbox User's Guide R2012b*. MathWorks, 2012.
- [235] V. Lemaire, F. Clérot, S. Busson, and V. Choqueuse. Individualized HRTFs from few measurements: a statistical learning approach. In *Proceedings of International Joint Conference on Neural Networks*, pages 2041–2046, Montreal, Canada, 2005.

-
- [236] Patricia G. Stelmachowicz, Kathryn A. Beauchaine, Ann Kalberer, and Walt Jesteadt. Normative thresholds in the 8- to 20-khz range as a function of age. *J. Acoust. Soc. Am.*, 86(4):1384–1391, 1989.
- [237] Song Xu, Zhizhong Li, and Gavriel Salvendy. Identification of anthropometric measurements for individualization of head-related transfer functions. *Acta Acustica united with Acustica*, 95(1):168–177, 2009.
- [238] David Schönstein and Brian FG Katz. Hrtf selection for binaural synthesis from a database using morphological parameters. In *20th International Congress on Acoustics, ICA*, 2010.
-