TECNIACUSTICA ELCHE · 2022

53º CONGRESO ESPAÑOL DE ACÚSTICA
XII CONGRESO IBÉRICO DE ACÚSTICA

OBTENCIÓN DE MODELOS 3D DE LA CABEZA Y OREJAS MEDIANTE CÁMARAS DE PROFUNDIDAD PARA LA PERSONALIZACIÓN DE LA HRTF

PACS: 43.66.Pn

Alvarez Martínez, Ariel; Universitat Politècnica de València, Camí de Vera, s/n, 46022**,** Valencia, Valencia, aalvmar@doctor.upv.es

López Monfort, José Javier; Universitat Politècnica de València, Camí de Vera, s/n, 46022**,** Valencia, Valencia, jjlopez@dcom.upv.es

ABSTRACT

The head-related transfer function (HRTF) describes how a human receives sound from different directions in space. It is unique to each listener and is essential for accurate virtual acoustic reproduction. The individualization of the HRTF can make an important contribution to improving the quality of binaural applications. Due to the advancement of technologies such as virtual reality, cinema and video games, the personalization of the HRTF has become a target of great interest. The best way to obtain an individualized HRTF is through direct measurement, but it requires sophisticated laboratory equipment and long measurement times, therefore several alternative methods have been explored, such as anthropometric models, physical synthesis and, more recently, deep learning models. In this work, anthropometric models are explored with the novelty of obtaining them through 3D models of the ear and head obtained through depth cameras (z axis) that are increasingly common in smartphones. The techniques used, the solution to the problems that have been appearing, as well as the first results of this method are described.

Keywords: HRTF, individualization, measurement, anthropometric models, 3D models.

RESUMEN

La función de transferencia relacionada con la cabeza (HRTF), describe cómo un humano recibe el sonido desde las diferentes direcciones del espacio. Es única para cada oyente y es imprescindible para una reproducción acústica virtual precisa. La individualización de la HRTF puede contribuir de manera importante a mejorar la calidad de las aplicaciones binaurales. Debido al avance de tecnologías como la realidad virtual, el cine y los videojuegos, la personalización de la HRTF se ha convertido en objetivo de gran interés. La mejor forma de obtener una HRTF individualizada es a través de la medición directa, pero requiere un equipo de laboratorio sofisticado y largos tiempos de medida, por lo que se han explorado varios métodos alternativos, tales como modelos antropométricos, síntesis física y, más recientemente, modelos de aprendizaje profundo. En este trabajo se exploran los modelos antropométricos con la novedad de obtenerlos mediante modelos 3D de la oreja y la cabeza obtenidos mediante las cámaras de profundidad (eje z) que cada vez son más habituales en los *smartphones*. Se describen las técnicas empleadas, la solución a los problemas que han ido apareciendo, así como los primeros resultados de este método.

Palabras Clave: HRTF, individualización, medida, modelos antropométricos, modelos 3D.

1.INTRODUCCIÓN

La función de transferencia relacionada con la cabeza (HRTF), describe cómo un humano recibe el sonido desde las diferentes direcciones del espacio [1]. Es única para cada oyente y es imprescindible para una reproducción acústica virtual precisa [2]. Debido a su singularidad, el uso de una HRTF genérica para aplicaciones acústicas a menudo genera errores de localización y una limitada percepción espacial [3]. Los estudios han demostrado que la individualización de la HRTF puede mejorar la precisión de la localización y las experiencias inmersivas de los usuarios [4,5]. Debido al avance de tecnologías como la realidad virtual, el cine y los videojuegos la personalización de la HRTF se ha convertido en un objetivo de gran interés. La mejor forma de obtener una HRTF individualizada es a través de la medición directa, pero es una tarea que requiere un equipo de laboratorio sofisticado y largos tiempos de medida, por lo que en los últimos años se han explorado otras alternativas menos invasivas. Estos métodos se pueden dividir en simulaciones numéricas y modelos antropométricos, el primer grupo consiste en simular la propagación de ondas acústicas alrededor del sujeto, los esquemas de simulación más comunes incluyen el método Fast Multipole Accelerated Boundary Method (FM-BEM) [6] y el método Finite Difference Time Domain Method (FDTD) [7]. Con la ayuda de las bases de datos de HRTFs disponibles públicamente y técnicas de aprendizaje automático, las mediciones antropométricas se pueden usar para elegir, adaptar o estimar el la HRTF de un sujeto. Por ejemplo, en 2010 Zeng et al. [8] implementó un modelo híbrido basado en el análisis de componentes principales (PCA) y la regresión lineal múltiple, que utilizó parámetros antropométricos para seleccionar el conjunto de HRTF más adecuado para un usuario determinado. Recientemente ha habido un creciente interés en resolver estas tareas con aprendizaje profundo, en 2017, Yao et al. [9] utilizó medidas antropométricas para seleccionar la HRTF más adecuada de una base de datos más grande, mediante el uso de redes neuronales. En 2018, Lee y Kim [10] desarrollaron una red neuronal de dos ramas que procesa datos antropométricos con un perceptrón multicapa (MLP) e imágenes del oído con un detector de bordes mediante capas convolucionales, combinando las salidas en una tercera red para estimar la HRTF de cada individuo.

Una parte esencial de los modelos antropométricos es la obtención de las medidas de la cabeza y orejas de cada individuo [11], la medición directa es una tarea compleja por las características de las orejas humanas principalmente, por este motivo se han utilizado técnicas de fotogrametría, visión artificial y aprendizaje automático para mediante imágenes 2D o escáneres 3D obtener las medidas antropométricas de cada individuo.

En este trabajo se explora la obtención de modelos 3D de cabeza y orejas mediante el uso de dispositivos equipados con cámaras de profundidad, se describen las técnicas, los materiales y los métodos empleados, así como los primeros resultados obtenidos.

2.MATERIALES Y MÉTODOS

Para el estudio se ha utilizado un smartphone (Iphone 13 Pro Max) con una aplicación para la toma de escáneres 3D (Heges 3D Scanner, de Apple Store, desarrollador: Marek Simonik) (Figura 1). El equipo cuenta con cámaras TrueDepth las cuales mediante un proyector de puntos brindan datos de profundidad en tiempo real junto con información visual, el sistema utiliza LEDs para proyectar una rejilla irregular de más de 30000 puntos infrarrojos que registran la profundidad en cuestión de milisegundos. La aplicación para la toma de escáneres 3D usada puede convertir las imágenes tomadas por las cámaras TrueDepth en modelos 3D con una resolución de 0.5 mm a 8.0 mm. Los modelos 3D generados pueden ser exportados en formato STL (sin color) o PLY (con color).

Durante la obtención del modelo 3D el sujeto fue sentado en una silla mirando hacia delante, manteniendo una posición natural de la cabeza. Se le colocó un gorro de natación para reducir la influencia del pelo en los escáneres y se le pidió que mantuviese una posición estática. Debido a la dificultad de obtener un escáner de la cabeza y orejas con la máxima precisión de 0.5 mm rotando el dispositivo 360 grados alrededor del sujeto, se decidió tomar 3 escáneres distintos, primeramente, se tomó uno de la cabeza completa, rotando el dispositivo 360 grados comenzando por la parte frontal de la cabeza, pero con una menor precisión de 1.0 mm (Figura 2), luego, se tomó uno por cada lado de la cabeza priorizando la zona de interés de las orejas,

con una precisión máxima de 0.5 mm, rotando el dispositivo 180 grados comenzando en la parte frontal de la cabeza y terminando en la parte posterior (Figuras 3 y 4). Todos los escáneres fueron tomados por un operario sosteniendo el dispositivo en la mano a una distancia del sujeto de 20 cm (Figura 1).

Figura 1 – Escáner usando la aplicación Heges 3D Scanner.

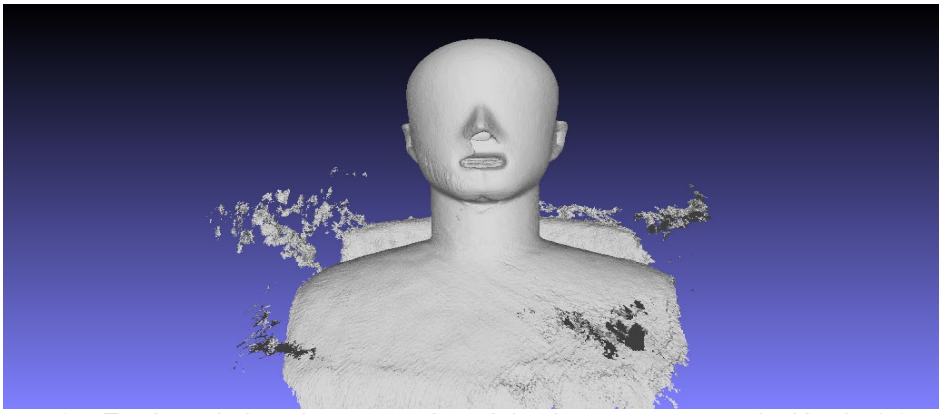

Figura 2 – Escáner de la cabeza completa del sujeto con una resolución de 1.0 mm.

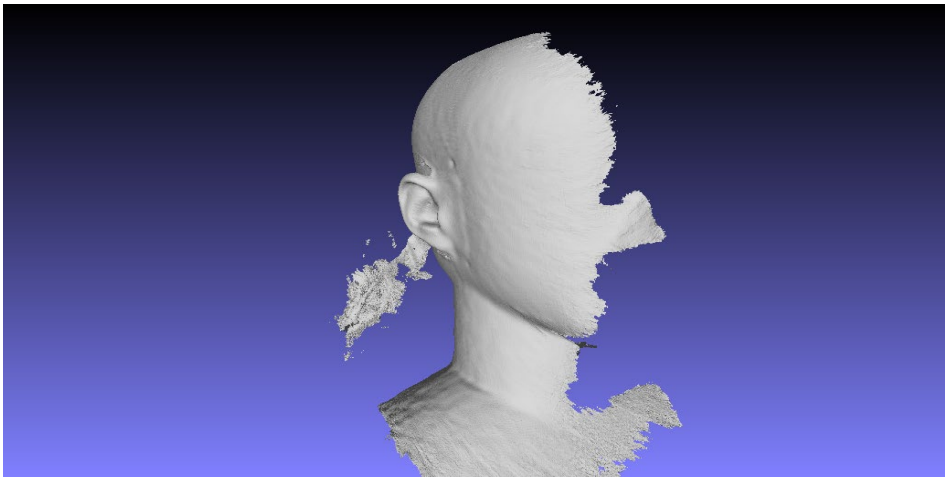

Figura 3 – Escáner del lado derecho del sujeto con una resolución de 0.5 mm.

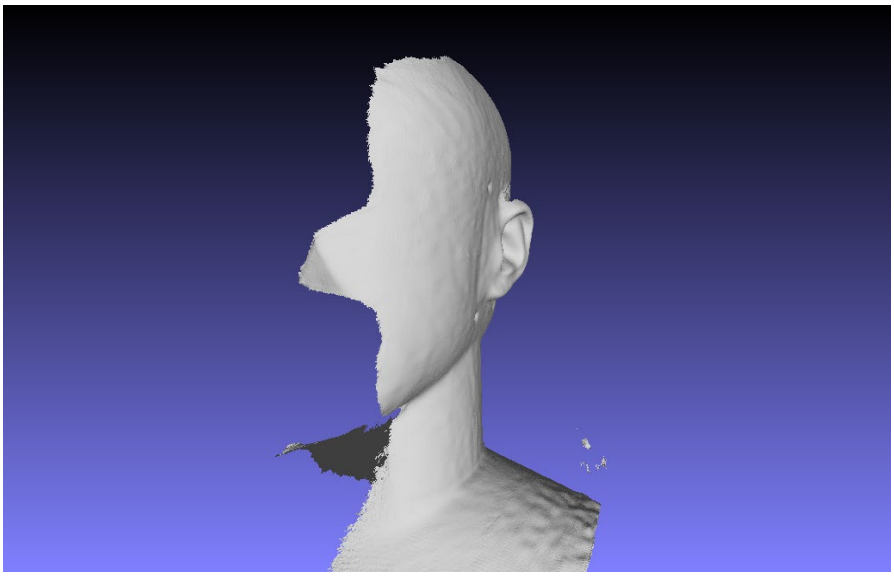

Figura 4 – Escáner del lado izquierdo del sujeto con una resolución de 0.5 mm.

Se realizaron varios escáneres en un tiempo aproximado de una hora, los tres mejores fueron seleccionados y exportados en formato STL. Las partes no deseadas fueron eliminadas usando MeshLab v2021.07. Para la obtención del modelo 3D final, en primer lugar, se eliminó las orejas en el escáner de la cabeza completa y luego se procedió a unir los tres escáneres utilizando la herramienta Point Based Glueing de MeshLab v2021.07. Los huecos del modelo final fueron cerrados usando Blender v2.93.7 (Figura 5). Finalmente, los modelos se alinearon al sistema de coordenadas basado en el eje interaural, definido como el eje que conecta los centros de las entradas a los canales auditivos. La alineación se realizó de forma semiautomática utilizando un script en Python v3.9.2 para Blender v2.93.7, que requería la selección de tres puntos en el modelo 3D (centro del canal auditivo izquierdo/derecho y un punto en la nariz).

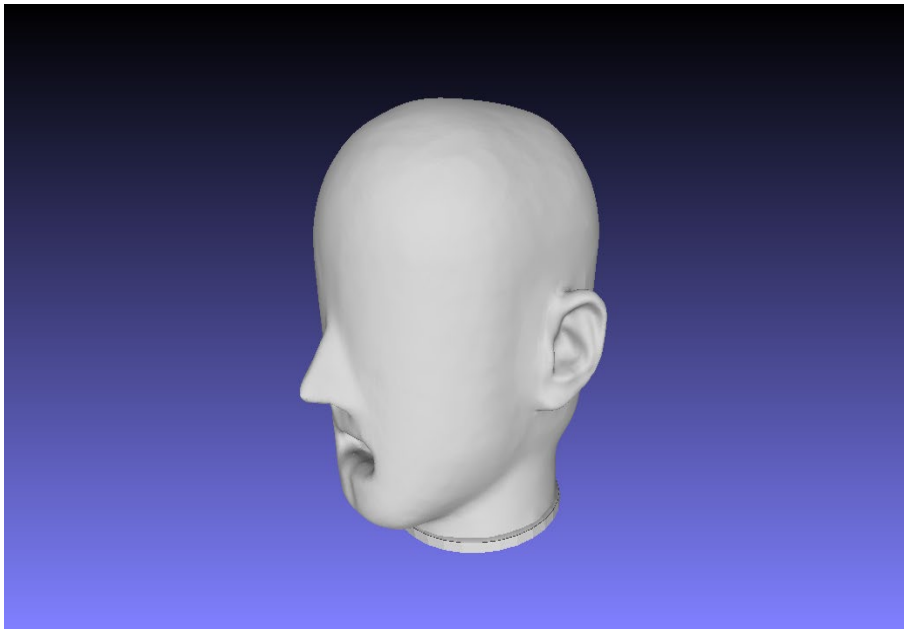

Figura 5 – Modelo final.

Debido a la influencia de la forma de las orejas en la HRTF principalmente en las altas frecuencias, se eligió el largo de las orejas como medida para comprobar la precisión del modelo 3D. Se utilizó la herramienta Measuring Tool de MeshLab v2021.07 para la obtención de la medida sobre el modelo 3D (Figura 6) y como referencia se utilizó la medición directa sobre el sujeto (Figura 7).

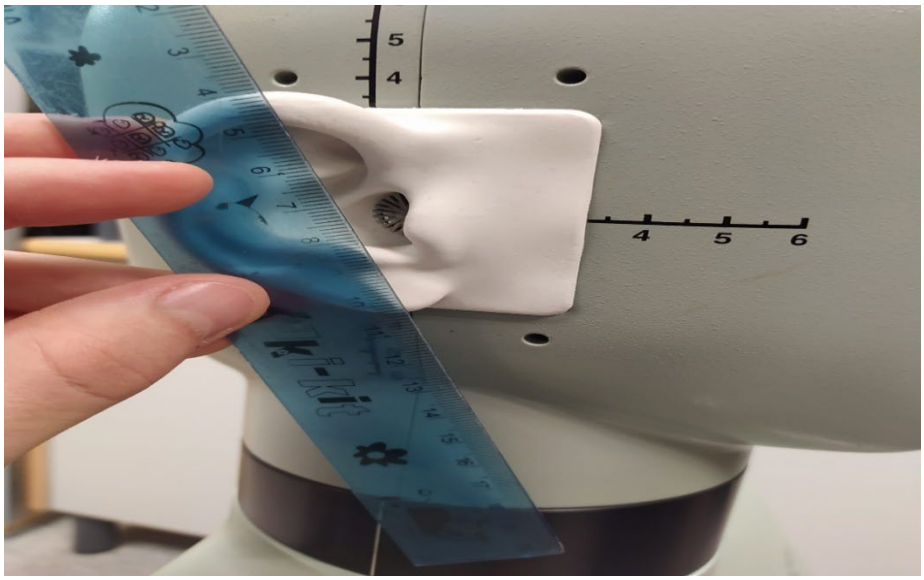

Figura 6 – Medida directa.

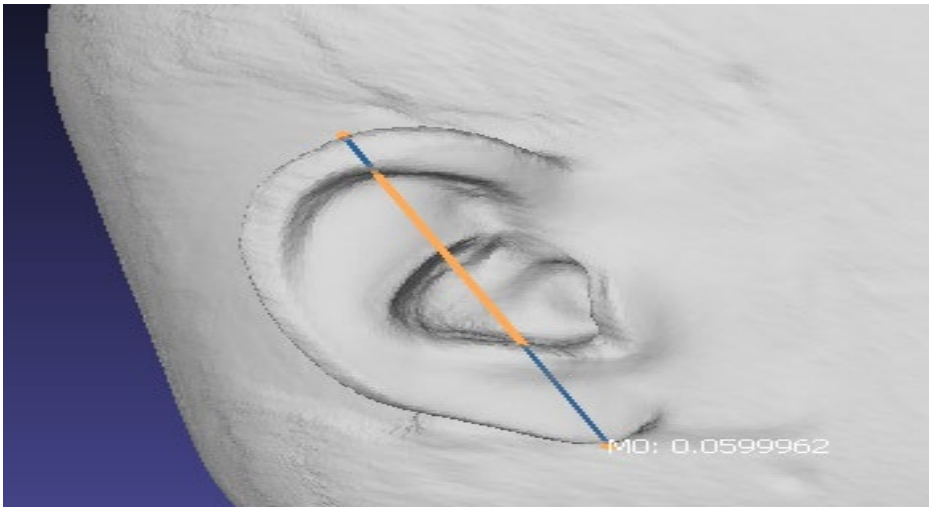

Figura 7 – Medida sobre el modelo.

La medida directa arrojó una distancia de aproximadamente 60.5 mm, mientras que sobre el modelo 3D se realizaron 10 medidas por la dificultad de seleccionar los puntos correctos, el promedio obtenido fue de 59.96 mm, por lo tanto, el error fue aproximadamente 0.89 %, un resultado positivo al encontrarse por debajo de 1%.

3. CONCLUSIONES

En este trabajo se ha propuesto un método para la obtención de modelos 3D de la cabeza y orejas, mediante el uso de smartphones equipados con cámaras de profundidad. El modelo final se logró a través de la unión de tres diferentes escáneres de la cabeza y orejas del sujeto, para conseguir este objetivo se utilizaron principalmente las herramientas MeshLab v2021.07 y Blender v2.93. Finalmente se comprobó la precisión del modelo mediante la herramienta Measuring Tool de MeshLab v2021.07, usando como medida de referencia la medición directa sobre el sujeto, se obtuvo un error de aproximadamente 0.89 %.

Los modelos 3D obtenidos serán utilizados en el futuro en sistemas antropométricos basados en visión artificial y aprendizaje automático, con el objetivo de extraer medidas de cada sujeto prescindiendo de las tediosas y largas medidas directas. Finalmente, se pretende utilizar las características antropométricas para alimentar redes neuronales que puedan generar en su salida la HRTF individualizada de cada sujeto.

AGRADECIMIENTOS

Este trabajo se enmarca en el proyecto “Intelligent Spatial Audio: machine-learning-assisted synthesis and montoring (ISLA)” (Referencia: RTI2018-097045-B-C22), financiado por la Unión Europea, el Gobierno de España y el Ministerio de Ciencia e Innovación a través de las ayudas para contratos predoctorales para la formación de doctores de 2019 (Referencia: PRE2019-089858).

Agradecemos al personal del Grupo de Tratamiento de Audio y Comunicaciones (GTAC) perteneciente al Instituto de Telecomunicaciones y Aplicaciones Multimedia (iTEAM) de la Universitat Politècnica de València, especialmente al Dr. Pablo Gutiérrez Parera por su tiempo, cariño y conocimientos. Al Dr. José Manuel Mossi García por su ayuda con la aplicación para la toma de escáneres.

REFERENCIAS

[1] Xie, B., Head-related transfer function and virtual auditory display, J. Ross Publishing, 2013.

[2] Kulkarni, A. and Colburn, H. S., “Role of spec- tral detail in sound-source localization,” Nature, 396(6713), p. 747, 1998.

[3] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? J. Audio Eng. Soc., 44(6):451–469, June 1996.

[4] Hu, H., Zhou, L., Ma, H., and Wu, Z., “HRTF per- sonalization based on artificial neural network in individual virtual auditory space,” Applied Acous- tics, 69(2), pp. 163–172, 2008.

[5] Armstrong, C., Thresh, L., Murphy, D., and Kear- ney, G., “A perceptual evaluation of individual and non-individual HRTFs: a case study of the SADIE II database,” Applied Sciences, 8(11), p. 2029, 2018.

[6] N. A. Gumerov, A. E. O’Donovan, R. Duraiswami, and D. N. Zotkin. Computation of the head-related transfer function via the fast multi- pole accelerated boundary element method and its spherical harmonic representation. J. Acoust. Soc. Am., 127(1):370–386, January 2010.

[7] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida. Mecha- nism for generating peaks and notches of head-related transfer func- tions in the median plane. J. Acoust. Soc. Am., 132(6):3832–3841, December 2012.

[8] X.-Y. Zeng, S.-G. Wang, and L.-P. Gao. A hybrid algorithm for select- ing head-related transfer function based on similarity of anthropometric structures. J. Sound Vibr., 329(19):4093–4106, Sept. 2010.

[9] S.-H. Yao, T. Collins, and C. Liang. Head-related transfer function selection using neural networks. Arch. Acoust., 42(3):365–373, Sept. 2017.

[10] G. W. Lee and H. K. Kim. Personalized HRTF modeling based on deep neural network using anthropometric measurements and images of the ear. Appl. Sci., 8(11), Nov. 2018.

[11] Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. The CIPIC HRTF database. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 99–102. 2001