

## ESTUDIO COMPARATIVO ENTRE REGRESIÓN LINEAL MÚLTIPLE Y REDES NEURONALES ARTIFICIALES PARA PERSONALIZAR HRTFS

**Fabián C. Tommasini, Oscar A. Ramos, Mariano Araneda, Sebastián P. Ferreyra y  
Gabriel A. Cravero**

*Centro de Investigación y Transferencia en Acústica, Universidad Tecnológica Nacional, Facultad  
Regional Córdoba. Córdoba, Argentina, [acustica@scdt.frc.utn.edu.ar](mailto:acustica@scdt.frc.utn.edu.ar);  
<http://www.investigacion.frc.utn.edu.ar/cintra>*

**Palabras Clave:** Acústica virtual, personalización de HRTFs, regresión lineal múltiple, redes neuronales artificiales.

**Resumen.** En realidad acústica virtual es crucial modelar al oyente mediante sus respuestas impulsivas de cabeza (HRIRs) o sus funciones de transferencia de cabeza (HRTFs). Medir las HRTFs a cada uno de los usuarios de un sistema de realidad acústica virtual es complejo, costoso y requiere de equipamiento y personal especializado. Por lo tanto, es necesario y conveniente desarrollar procedimientos más simples y baratos que permitan estimar las HRTFs de un sujeto sin tener que medirlas. Las ondas sonoras sufren complejas transformaciones al interactuar con la cabeza, el torso y la oreja, antes de alcanzar los tímpanos del oyente. Uno de los abordajes más utilizados para personalizar HRTFs es el de apelar a mediciones de algunas características antropométricas del individuo, responsables de las transformaciones mencionadas. Para construir HRTFs personalizadas, es común utilizar el análisis de componentes principales (PCA) aplicado a un conjunto de HRTFs medidas a un grupo de sujetos. Luego, se busca establecer la relación entre parámetros antropométricos seleccionados y los pesos del PCA, utilizando algún modelo de regresión multivariable. El propósito de este trabajo fue contrastar las HRTFs obtenidas para sujetos no incluidos en el grupo anterior, mediante dos métodos: regresión lineal múltiple y redes neuronales artificiales (ANNs). Se utilizó la distancia euclidiana entre los dos espectros para estimar la calidad del ajuste entre las HRTFs medidas y las estimadas. El estudio fue realizado en el plano medio en el cual los perfiles espectrales son esenciales para localizar una fuente sonora. Los resultados mostraron que los errores de las ANNs son menores para la mayoría de las posiciones.

## 1. INTRODUCCIÓN

El propósito de la realidad acústica virtual (acoustic virtual reality, AVR) es recrear la experiencia auditiva que una persona experimentaría en un entorno real, provocando la sensación de presencia (inmersión) en dicho entorno. El principio que sustenta la simulación acústica establece que: estímulos acústicos equivalentes, evocan sensaciones equivalentes (Blauert, 1997). En otras palabras, si mediante auriculares se aplican a los tímpanos de un oyente las señales biológicamente correctas, será posible estimular una sensación de inmersión en el entorno modelado (Kleiner et al., 1993).

En un sistema de AVR se deben modelar tres grandes subsistemas: la fuente sonora, el recinto y el oyente. La fuente sonora se caracteriza por su directividad y su respuesta en frecuencia; el recinto, por su respuesta impulsiva para una posición específica de la dupla fuente-oyente; y el oyente, por sus respuestas impulsivas de cabeza (head-related impulse responses, HRIRs) en el dominio del tiempo, o sus funciones de transferencia de cabeza (head-related transfer functions, HRTFs) en el dominio de la frecuencia (Tommasini, 2012).

Cada HRIR o HRTF<sup>1</sup> contiene las transformaciones que se producen en una onda sonora antes de alcanzar los tímpanos de un oyente, al interactuar con el torso, los hombros y especialmente con la cabeza y los pabellones auriculares. Son diferentes para cada oído y varían sistemáticamente con la ubicación de la fuente sonora en el espacio. Las HRTFs son medidas entre una fuente sonora ubicada en una superficie esférica y micrófonos miniaturas colocados en la proximidad de los tímpanos o a la entrada de los conductos auditivos.

Una HRTF también contiene información sobre el oyente, es decir, varía también con cada sujeto. Esto es debido a las diferencias antropométricas tales como: tamaño de la cabeza y orejas; ancho de hombro y torso; entre otras. Si el conjunto de HRTFs utilizadas corresponde a las del oyente, la fuente se percibe compacta, externa y bien definida en una posición del espacio. Por el contrario, si pertenecen a otro individuo, la fuente se percibe difusa, ubicada en el interior de la cabeza y existen confusiones delante-detrás (Wenzel et al., 1993; Blauert, 1997). Esto implicaría medir a cada oyente sus propias HRTFs. Sin embargo, estas mediciones son muy complejas de realizar y se necesita de equipamiento especial, lo que es muy difícil y caro en la práctica. Por esto es importante desarrollar métodos que permitan estimar las HRTFs de un individuo a partir de mediciones más simples y baratas.

Varios estudios han abordado el problema de personalización de HRTFs de diferentes formas. Una revisión se puede consultar en Xu et al. (2007), donde se identifican siete métodos potenciales para personalizar HRTFs. Algunos métodos son el *database matching* (Zotkin et al., 2003), modificaciones del espectro (Middlebrooks, 1999) y el método de elementos de contorno (boundary element method, BEM) (Otani y Ise, 2000). Uno de los métodos más difundidos se basa en obtener la relación entre algunas mediciones antropométricas del individuo y la magnitud logarítmica de las HRTFs mediante regresión lineal múltiple, por ejemplo, Rodríguez y Ramírez (2005); Hu et al. (2006); Hwang y Park (2007); Xu et al. (2009); Ramos et al. (2009). Debido al hecho de que la influencia de la cabeza, la oreja y el torso tiene características complejas, otros autores utilizan métodos de regresión no lineal, por ejemplo, Hu et al. (2008); Huang y Fang (2009).

Para evaluar objetivamente la calidad de ajuste entre las HRTFs medidas y las estimadas (personalizadas), se utiliza habitualmente la distancia euclidiana entre los dos espectros, integrado en todo el rango de frecuencia, conocida como distorsión espectral (spectral distortion, SD) (Nishino et al., 1999).

<sup>1</sup>En adelante se las nombrará de forma indistinta, salvo que se explicita el dominio utilizado.

El propósito de este trabajo es contrastar HRTFs personalizadas en el plano medio mediante dos métodos de regresión, uno lineal, como es la regresión lineal múltiple, y el otro no lineal, como las redes neuronales artificiales (artificial neural networks, ANNs). En la Sección 2 se explica la metodología seguida: la base de datos de HRIRs utilizada, el pre-procesamiento aplicado y la selección de los parámetros antropométricos relevantes. La Sección 3 describe los métodos utilizados: regresión lineal múltiple y ANNs. En la Sección 4 se muestran los resultados obtenidos. Finalmente, en la Sección 5 se presentan las conclusiones.

## 2. METODOLOGÍA

### 2.1. La base de datos

La base de datos utilizada es la provista por el Center for Image Processing and Integrated Computing (CIPIC) de la Universidad de California (Algazi et al., 2001). Contiene mediciones de HRIRs (dominio del tiempo) realizadas a 47 sujetos en 1250 posiciones (Figura 1). La ubicación de la fuente sonora se especifica por el ángulo de azimut  $\theta$  (25) y el ángulo de elevación  $\phi$  (50) referenciados a un eje que pasa por ambos oídos denominado eje inter-aural. Las HRIRs son secuencias de 200 puntos muestreados a una frecuencia de 44.1 kHz. Están compensadas en campo libre y medidas con la entrada del canal auditivo bloqueada.

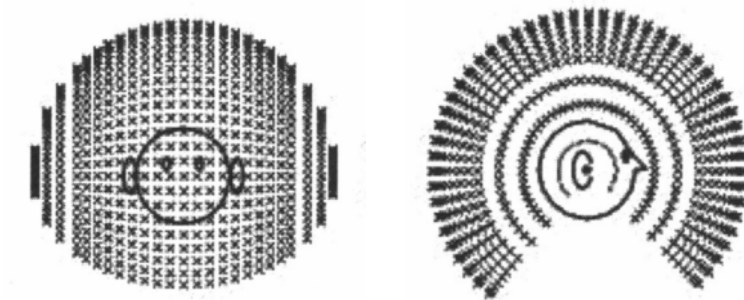


Figura 1: Representación gráfica de las 1250 posiciones (extraído de Algazi et al. (2001)).

Las mediciones se realizaron para elevaciones que varían uniformemente en escalones de  $5.625^\circ$  entre  $-45^\circ$  y  $+230.625^\circ$ . Los ángulos de azimut son:  $-80^\circ$ ,  $-65^\circ$ ,  $-55^\circ$ , desde  $-45^\circ$  a  $+45^\circ$  en escalones de  $5^\circ$ ,  $+55^\circ$ ,  $+65^\circ$  y  $+80^\circ$ . Azimut  $0^\circ$  y elevación  $0^\circ$ , corresponde al frente; azimut  $0^\circ$  y elevación  $+180^\circ$  detrás del sujeto; azimut negativos a la izquierda y positivos a la derecha del sujeto; elevaciones negativas y mayores a  $+180^\circ$  debajo del eje inter-aural al frente y detrás respectivamente.

En el presente trabajo se utilizarán las 50 posiciones del plano medio ( $\theta = 0^\circ$ ). Además, como es posible despreciar las asimetrías de la cabeza, se trabajará sobre el oído izquierdo.

Esta base de datos también incluye mediciones antropométricas para cada sujeto. En la Figura 2 se muestran las 17 mediciones de la cabeza y torso y las 10 en cada uno de los oídos. La totalidad de de estos parámetros fueron medidos a 35 de los 47 sujetos. Por lo tanto, sólo éstos fueron utilizados en este estudio.

### 2.2. Pre-procesamiento

Mehrgardt y Mellert (1977) demostraron que, en la mayoría de las posiciones, las HRTFs son secuencias de fase-mínima y que el resto de fase, esto es, la diferencia entre la fase total me-

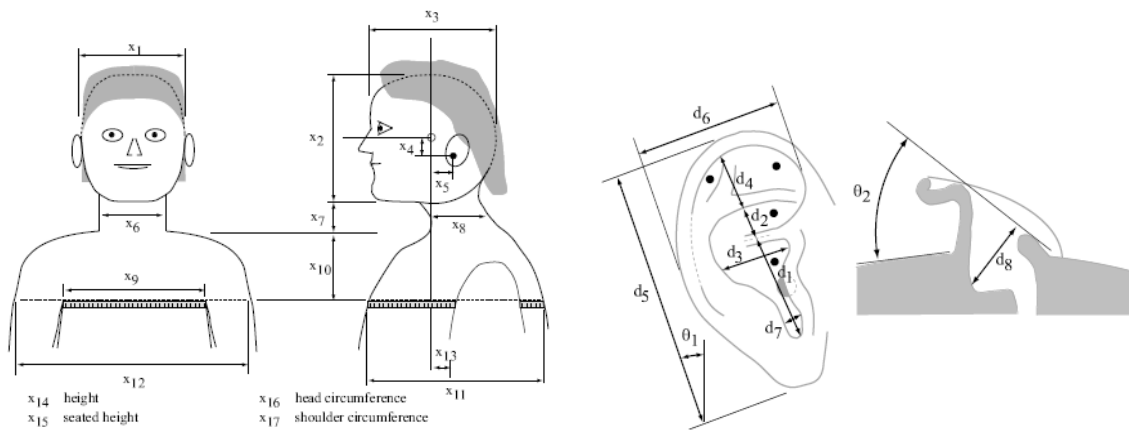


Figura 2: Mediciones de cabeza, torso y oído externo (extraído de [Algazi et al. \(2001\)](#)).

nos la fase mínima, es prácticamente lineal e independiente de la frecuencia. Estas evidencias permitieron desarrollar un modelo simplificado de las HRTFs conocido como *minimum-phase-plus-delay* (fase-mínima-más-retardo). La magnitud de las HRTFs puede ser aproximada por las secuencias de fase mínima y su fase por  $-\omega\tau$ , donde  $\tau$  es un retardo constante e independiente de la frecuencia. Pruebas psicoacústicas han corroborado que la habilidad de los sujetos para localizar una fuente sonora no se degrada si se utilizan HRTFs generadas por este modelo ([Kistler y Wightman, 1992](#)).

En primer lugar, se extrajeron los retardos inter-aurales (interaural delays, ITDs) de las HRIRs (dominio del tiempo) de la base de datos del CIPIC. Luego, se obtuvieron las HRTFs (dominio de la frecuencia) correspondientes, realizando la transformada discreta de Fourier (completando con ceros hasta 256 muestras). Como el ancho de banda bajo estudio fue de 0–15 kHz ([Kistler y Wightman, 1992](#)), se seleccionaron  $N = 88$  componentes de frecuencia (desde 0 hasta 14.987 kHz, con una resolución cercana a los 172 Hz).

Luego, de los 35 sujetos, se separó un grupo de  $Q = 5$  sujetos al azar (llamado S5) que permitirá validar las HRTFs personalizadas por ambos métodos. Al grupo de los  $P = 30$  sujetos restantes (llamado S30), se le calculó la media  $\mu_h$  de la magnitud logarítmica de las HRTFs para todos los sujetos, posiciones y ambos oídos ([Kistler y Wightman, 1992](#)). Esta media se restó de cada HRTF para eliminar las características comunes del grupo S30, y se obtuvieron las HRTFs pre-procesadas  $\mathbf{h}_p$ , para un sujeto  $p$ . Luego, se aplicó un análisis de componentes principales (principal component analysis, PCA) a este grupo ([Martens, 1987; Kistler y Wightman, 1992](#)). El PCA transforma un número de variables correlacionadas en un mismo número de variables no correlacionadas, conocidas como componentes principales (principal components, PCs). Cada PC captura, en mayor o menor grado, algún aspecto común del conjunto. Además, permite reducir la dimensionalidad del conjunto.

De la aplicación del PCA se obtuvo una matriz  $N \times N$  de transformación ortonormal  $\mathbf{C}$  que contiene las PCs y una matriz  $P \times N$  de pesos  $\mathbf{W}$  para cada posición y para cada oído. Cualquier HRTF para un sujeto  $p$  y para un posición específica  $(\theta, \phi)$  puede ser completamente reconstruída mediante

$$\mathbf{h}_p(\theta, \phi) = \mathbf{w}_p(\theta, \phi)\mathbf{C} + \mu_h \quad \text{con } p = 1 \dots P \quad (1)$$

donde  $\mathbf{h}_p$  es un vector de longitud  $N$  el cual contiene la HRTF de cualquier oído del sujeto  $p$  perteneciente a S30,  $\mathbf{w}_p$  es el  $p$ -ésimo vector fila de la matriz  $\mathbf{W}$  y  $\mu_h$  es la media calculada

anteriormente.

Dada que las PCs derivadas del grupo S30 se pueden considerar como PCs genéricas (Hwang et al., 2008), fueron utilizadas para obtener otra matriz  $Q \times N$  de pesos  $\mathbf{W}'$  para cada posición y para cada oído para el grupo S5. A partir de la Ecuación (1), se puede escribir

$$\mathbf{w}'_q = (\mathbf{h}_q(\theta, \phi) - \mu_{\mathbf{h}}) \mathbf{C}^T \quad \text{con } q = 1 \dots Q \quad (2)$$

donde  $\mathbf{h}_q$  contiene la HRTF para cada oído del sujeto  $q$  perteneciente a S5,  $\mathbf{w}'_q$  es el  $q$ -ésimo vector fila de la matriz  $\mathbf{W}'$ , y  $\mathbf{C}^T$  es la traspuesta<sup>2</sup> de  $\mathbf{C}$ .

Sin embargo, para obtener una disminución efectiva de la dimensión de los datos originales, se debe determinar un número  $L \ll N$  que permita reconstruir una HRTF dada dentro de un error perceptualmente aceptable (Ramos et al., 2007; Ramos y Tommasini, 2008). De acuerdo a Ramos et al. (2010) las 4 primeras PCs reflejan la influencia del oído externo y la cabeza, y las reflexiones producidas en torso y hombros influyen a partir de la PC 5 hasta la 18. Como en el plano medio la influencia del oído externo es preponderante, se utilizaron  $L = 4$  PCs (que representa el 85.09 % de la variación de los datos).

Las HRTFs reconstruídas con la Ecuación (1) para el grupo S5 usando  $\mathbf{W}'$  con  $L = 4$  PCs serán referenciadas como  $\text{HRTF}_{\text{ori}}$ .

### 2.3. Selección de parámetros antropométricos

Utilizando la base de datos de HRIRs del CIPIC y apelando al método del PCA, numerosos autores se propusieron, por un lado, determinar la conexión entre las dimensiones de los rasgos antropométricos involucrados y la estructura de las HRIRs y, por el otro, construir HRTF personalizadas (Hu et al., 2008; Ramos y Tommasini, 2008; Hwang y Park, 2007, 2008; Xu et al., 2009). Las metodologías de selección de las mediciones antropométricas relevantes son disímiles y se aborda el tema casi exclusivamente como un problema matemático. La selección de los parámetros antropométricos se basa en asumir índices de correlación entre los parámetros antropométricos y las HRTFs mayores a cierto valor adoptado arbitrariamente. Como consecuencia de este abordaje, quedan excluidas medidas antropométricas que la literatura ha demostrado su influencia en la composición de las HRTFs. Por ejemplo: una correlación significativa entre la altura de un sujeto y el ancho de la cabeza, no amerita descartar esa medición de la cabeza que, como se sabe, su influencia en las HRTFs es decisiva. Además, llama la atención que, utilizando la misma base de datos, de los 13 parámetros antropométricos seleccionados por Hu et al. (2008) y de los 10 por Xu et al. (2009) sólo tienen en común tres de ellos.

En un trabajo reciente (Fels y Vorländer, 2009) se estudió la influencia de las dimensiones de la cabeza, torso y oído externo en las HRTFs de individuos que van desde la niñez (6 meses) a la adultez (23 años). Su objetivo fue conocer cuáles son de estas dimensiones son las que determinan los detalles espectrales de las HRTFs y sus correspondientes claves binaurales. Para el presente trabajo se utilizaron las medidas del oído externo basadas en este estudio, ya que en el plano medio su influencia es preponderante. Los parámetros utilizados fueron: el alto y ancho de las orejas ( $d_5$  y  $d_6$ ); el alto, ancho y profundidad de la concha ( $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_8$ ); y el ángulo de rotación de la oreja ( $\theta_1$ ). En resumen, los  $K = 7$  parámetros antropométricos seleccionados fueron:  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_5$ ,  $d_6$ ,  $d_8$  y  $\theta_1$ .

<sup>2</sup> $\mathbf{C}^T = \mathbf{C}^{-1}$  dado que la matriz  $\mathbf{C}$  es ortonormal.

### 3. MODELADO POR REGRESIÓN

Se utilizaron dos métodos de regresión: uno lineal, como es la regresión lineal múltiple, y el otro no lineal, como las ANNs. Para luego evaluar cuál tiene un menor error en la personalización de HRTFs. Una vez estimada la HRTF personalizada para una posición y para un sujeto, se utiliza el modelo de fase-mínima-más-retardo, es decir, se agrega el ITD correspondiente a la reconstrucción final de la HRIR (dominio del tiempo).

Tanto el pre-procesamiento anteriormente descripto, como ambos métodos de regresión y el proceso de reconstrucción de una HRTF fueron implementados en el entorno MATLAB de MathWorks.

#### 3.1. Regresión lineal múltiple

El método consiste en expresar la matriz  $P \times L$  de pesos  $\mathbf{W}$  (de todos los sujetos del grupo S30) para una posición  $(\theta, \phi)$  de la fuente sonora, como una combinación lineal de las  $K = 7$  dimensiones antropométricas seleccionadas

$$\mathbf{W}(\theta, \phi) = \mathbf{X}\mathbf{B}(\theta, \phi) + \mathbf{E}(\theta, \phi) \quad (3)$$

donde  $\mathbf{X}$  es la matriz  $P \times K$  de parámetros antropométricos de los sujetos del grupo S30,  $\mathbf{B}$  es la matriz  $K \times L$  de coeficientes de regresión y  $\mathbf{E}$  la matriz  $P \times L$  de errores de estimación.

Luego, los coeficientes de regresión son calculados haciendo (Hu et al., 2006)

$$\mathbf{B}(\theta, \phi) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(\theta, \phi). \quad (4)$$

Con esta matriz de coeficientes de regresión  $\mathbf{B}$  y la matriz  $Q \times K$  de parámetros antropométricos  $\mathbf{Y}$  de los sujetos del grupo S5, se puede calcular una matriz  $Q \times L$  de pesos  $\hat{\mathbf{W}}_{\text{mlr}}$  para dichos sujetos

$$\hat{\mathbf{W}}_{\text{mlr}}(\theta, \phi) = \mathbf{Y}\mathbf{B}(\theta, \phi). \quad (5)$$

Usando estos pesos estimados y la Ecuación (1), se pueden reconstruir las HRTFs para cualquier sujeto del grupo S5. A estas HRTFs se las llamará  $\text{HRTF}_{\text{mlr}}$ .

#### 3.2. Redes neuronales artificiales

Las redes neuronales artificiales (ANNs) son sistemas de procesamiento inspirados en el cerebro humano y, dado que una de sus principales características es su capacidad de aprendizaje, han sido ampliamente utilizadas para describir relaciones no lineales entre entrada-salida en diferentes aplicaciones.

El tipo de ANN más utilizada para una función de regresión no lineal múltiple es una ANN *feedforward* con funciones de activación sigmoideas en su capa oculta, la cual es entrenada por el algoritmo de *backpropagation* (Wythoff, 1993). Este algoritmo es, básicamente, un descenso por la gradiente diseñado para minimizar la función de error en el espacio de pesos de la red. El teorema de Kolmogorov muestra que cualquier función continua puede ser aproximada uniformemente por el modelo de la red *backpropagation* con sólo una capa oculta (Hecht-Nielsen, 1987).

Dado que las componentes espectrales de una HRTF tienen una relación muy compleja con las características antropométricas, es difícil expresarla en una fórmula lineal. En este caso, se utilizaron varias ANNs para modelar la regresión no lineal entre los  $K = 7$  parámetros antropométricos seleccionados y los primeros  $L = 4$  pesos del PCA.



Se utilizan en total 50 ANNs, una para cada posición del plano medio. Cada una de estas ANNs tiene una arquitectura de tres capas, con funciones de activación tan-sigmoidales (Figura 3a) en la capa oculta y funciones de activación lineales (Figura 3b) en la capa de salida. La elección del número de neuronas en la capa oculta es muy importante pero, sin embargo, no existe una fórmula general para conocerlo a priori. Un número grande de neuronas en esta capa ofrece una mayor flexibilidad debido a que la red tiene más parámetros que puede optimizar. Pero si este número se hace demasiado grande, puede causar el problema de sub-caracterización y la red debe optimizar más parámetros que los necesarios. Para determinar este número se fue incrementando gradualmente la cantidad de neuronas ocultas de 5 a 25. El mínimo error se alcanzó con 15 neuronas ocultas. Por lo tanto, la estructura de cada ANN resultó ser de 7–15–4.

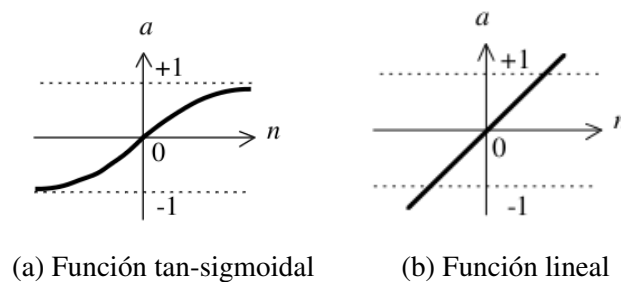


Figura 3: Funciones de activación de la ANN.

Para la fase de entrenamiento se utilizaron los parámetros antropométricos y los pesos  $w_p$  del grupo S30. Para la fase de *testing* se utilizaron los parámetros y los pesos  $w'_q$  del grupo S5. Tanto en la fase de entrenamiento como en la fase de *testing*, la entrada y la salida fueron normalizadas. Todas las ANNs fueron entrenadas mediante el método regularización Bayesiana (Bayesian regularization) hasta 200 iteraciones con un error objetivo de 0.02. Este método mejora la generalización y la medida del rendimiento es la suma de los errores cuadrados.

La estimación de los  $L = 4$  pesos en la salida de cada ANN, conforman la matriz  $\hat{W}_{ann}$  de tamaño  $Q \times L$ . Usando esta matriz de pesos y la Ecuación (1), se calcularon las HRTFs para el grupo S5. Estas HRTFs serán referenciadas como  $HRTF_{ann}$ .

#### 4. RESULTADOS

Para evaluar el nivel de ajuste, se reconstruyeron las HRTFs ( $HRTF_{ori}$ ,  $HRTF_{mlr}$  y  $HRTF_{ann}$ ) para el grupo de sujetos S5, los cuales no estaban incluidos en el cálculo del PCA. Esto difiere del artículo de referencia (Hu et al., 2008), que calcula los errores para los mismos sujetos involucrados en el cálculo del PCA.

Para evaluar los errores entre las  $HRTF_{ori}$  y las HRTFs calculadas ( $HRTF_{mlr}$  y  $HRTF_{ann}$ ) se utilizó la diferencia media cuadrática de la magnitud logarítmica de las HRTFs, es decir, la SD

$$SD(\theta, \phi) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( h(\theta, \phi, f_i) - \hat{h}(\theta, \phi, f_i) \right)^2} \quad [\text{dB}] \quad (6)$$

donde  $h$  es la componente en frecuencia de  $HRTF_{ori}$ ,  $\hat{h}$  es la estimada ( $HRTF_{mlr}$  y  $HRTF_{ann}$ ), y  $N = 88$  es el número de componentes de frecuencia que representan un rango de 0–15 kHz.

En la Figura 4a se muestran los valores SD promedio de todos los sujetos del grupo S5 para todas las posiciones del plano medio. En la Figura 4b se observan los valores SD más/menos

una desviación estándar en ambos métodos. Como se puede observar los errores de las ANNs son menores para la mayoría de las posiciones.

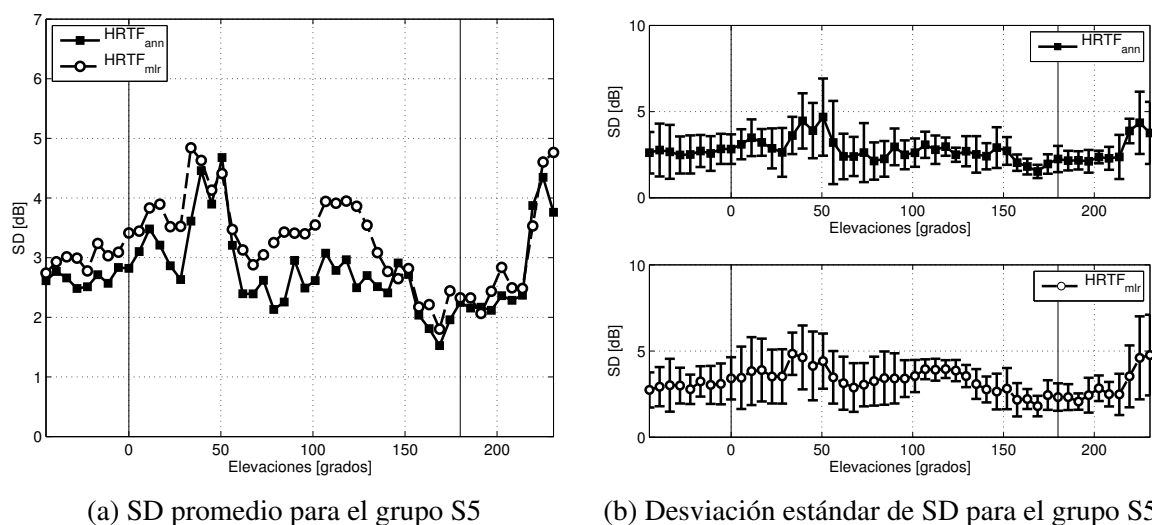


Figura 4: Valores de SD en el plano medio para ambos métodos en el rango 0–15 kHz.

Para verificar qué método es significativamente mejor, se realizó una prueba  $t$  de Student (con nivel de significancia  $\alpha = 0.01$ ) para el grupo S5. Dado que la diferencia promedio de los valores de SD de ambos métodos siguen una distribución normal, se plantearon las siguientes hipótesis

$$\begin{aligned} H_0 &: \mu_{\text{ann}} - \mu_{\text{mlr}} \geq 0 \\ H_1 &: \mu_{\text{ann}} - \mu_{\text{mlr}} < 0 \end{aligned} \quad (7)$$

donde  $\mu_{\text{ann}}$  es el valor SD promedio de las HRTF<sub>ann</sub> y  $\mu_{\text{mlr}}$  es el valor SD promedio para las HRTF<sub>mlr</sub>. Se debe notar que estas hipótesis constituyen una prueba de una cola y la hipótesis nula  $H_0$  será rechazada si  $\mu_{\text{ann}} < \mu_{\text{mlr}}$ .

Como  $H_0$  fue rechazada ( $t = -8.3485$ ,  $p < 0.01$ ), indica que el método de las ANNs tienen, en promedio, un valor SD significativamente menor que el método de la regresión lineal múltiple para los sujetos del grupo S5.

## 5. CONCLUSIONES

Se utilizó el PCA para reducir la dimensionalidad de un conjunto de HRTFs medidas a un grupo de 30 sujetos. Luego, con el objetivo de personalizar HRTFs, se estimaron los pesos del PCA para dicho grupo mediante dos métodos: regresión lineal múltiple y ANNs, usando 7 parámetros antropométricos seleccionados. Estos parámetros son mediciones del oído externo basadas en el estudio de Fels y Vorländer (2009). La estimación de las HRTFs personalizadas con 4 PCs fueron comparadas con las HRTFs reconstruidas de 5 sujetos externos al grupo previo.

En este artículo se trabajó sobre el plano medio, que es donde existen los mayores errores en la personalización de las HRTFs. La mayoría de los estudios de referencia trabajan en el plano horizontal donde los errores son menores.

A partir de los resultados obtenidos se puede concluir que las ANNs son un mejor método para personalizar HRTFs en el plano medio que la regresión lineal múltiple. Si bien los resulta-



dos parecen ser promisorios, en futuros trabajos, se deberá realizar pruebas psicoacústicas que permitan validarlos perceptualmente.

## REFERENCIAS

- Algazi V.R., Duda R.O., Thompson D.M., y Avendano C. The CIPIC HRTF database. En *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, páginas 99–102. New Paltz, NY, USA, 2001.
- Blauert J. *Spatial hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, MA, USA, 1997.
- Fels J. y Vorländer M. Anthropometric parameters influencing head-related transfer functions. *Acta Acustica united with Acustica*, 95:331–342, 2009.
- Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem. En *Proceedings of the Second IEEE International Conference on Neural Networks*, volumen 3, páginas 11–13. 1987.
- Hu H., Zhou L., Ma H., y Wu Z. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, 69(2):163–172, 2008.
- Hu H., Zhou L., Zhang J., Ma H., y Wu Z. Head related transfer function personalization based on multiple regression analysis. En *Computational Intelligence and Security, 2006 International Conference on*, volumen 2, páginas 1829–1832. 2006.
- Huang Q.h. y Fang Y. Modeling personalized head-related impulse response using support vector regression. *Journal of Shanghai University (English Edition)*, 13(6):428–432, 2009.
- Hwang S. y Park Y. HRIR customization in the median plane via principal components analysis. En *Proceedings of AES 31st Int. Conf. New Directions in High Resolution Audio*. London, UK, 2007.
- Hwang S. y Park Y. Interpretations on principal components analysis of head-related impulse responses in the median plane. *The Journal of the Acoustical Society of America*, 123(4):EL65, 2008.
- Hwang S., Park Y., y Park Y.s. Modeling and customization of head-related impulse responses based on general basis functions in time domain. *Acta Acustica united with Acustica*, 94(6):965–980, 2008.
- Kistler D.J. y Wightman F.L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647, 1992.
- Kleiner M., Dalenbäck B., y Svensson U.P. Auralization-an overview. *Journal of the Audio Engineering Society*, 41(11):861–875, 1993.
- Martens W.L. Principal components analysis and resynthesis of spectral cues to perceived direction. En *Proceedings of the 1987 International Computer Music Conference*, páginas 274–281. 1987.
- Mehrgardt S. y Mellert V. Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America*, 61(6):1567–1576, 1977.
- Middlebrooks J.C. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, 106(3):1493–1510, 1999.
- Nishino T., Kajita S., Takeda K., y Itakura F. Interpolation of the head related transfer function on the horizontal plane. *Journal of the Acoustical Society of Japan*, 55:91–99, 1999.
- Otani M. y Ise S. Numerical calculation of the head-related transfer functions by using the boundary element method. En *Proceedings of WESTPRAC VII*, volumen 1, páginas 305–

308. 2000.

- Ramos O.A., Araneda M., y Tommasini F.C. Construcción de HRTF individuales a partir de mediciones antropométricas. *Mecánica Computacional*, XXVIII:149–161, 2009.
- Ramos O.A., Calvo G., y Tommasini F.C. Modelo acústico de cabeza y torso mediante análisis de componentes principales. *Mecánica Computacional*, XXVI:46–58, 2007.
- Ramos O.A. y Tommasini F.C. Simplificación de las funciones de transferencia de cabeza mediante análisis de componentes principales. *Mecánica Computacional*, XXVII:431–442, 2008.
- Ramos O.A., Tommasini F.C., y Araneda M. Contribución de la cabeza, el torso y el oído externo en las funciones de transferencia relativas a la cabeza. En *2do Congreso Internacional de Acústica UNTREF*. Buenos Aires, Argentina, 2010.
- Rodríguez S.G. y Ramírez M.A. Linear relationships between spectral characteristics and anthropometry of the external ear. En *Proceedings of 11 Meeting of the International Conference on Auditory Display*, páginas 336–339. Limerick, Irlanda, 2005.
- Tommasini F.C. *Sistema de simulación acústica virtual en tiempo real*. Doctoral thesis, Universidad Nacional de Córdoba. Facultad de Ciencias Exactas, Físicas y Naturales, Córdoba, Argentina, 2012.
- Wenzel E.M., Arruda M., Kistler D.J., y Wightman F.L. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993.
- Wythoff B.J. Backpropagation neural networks: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 18(2):115–155, 1993.
- Xu S., Li Z., y Salvendy G. Individualization of head-related transfer function for three-dimensional virtual auditory display: a review. En *Proceedings of the 2nd international conference on Virtual reality, ICVR'07*, páginas 397–407. Springer-Verlag, Berlin, Heidelberg, 2007.
- Xu S., Li Z., y Salvendy G. Identification of anthropometric measurements for individualization of Head-Related transfer functions. *Acta Acustica united with Acustica*, 95(1):168–177, 2009.
- Zotkin D.N., Hwang J., Duraiswami R., y Davis L.S. HRTF personalization using anthropometric measurements. En *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on*, páginas 157–160. 2003.