

PREPROCESAMIENTO EFICAZ PARA MODELAR LAS RESPUESTAS IMPULSIVAS DE CABEZA, UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES

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

*Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET,
Universidad Tecnológica Nacional – Facultad Regional Córdoba, Mtro .M .López esq .Cruz Roja
Argentina, Ciudad Universitaria, 5016 Córdoba, Argentina. oramos@scdt.frc.utn.edu.ar*

Palabras Clave: HRTF, análisis de componentes principales, realidad acústica virtual.

Resumen. Es muy frecuente, tanto en el dominio del tiempo como el de la frecuencia, el uso del análisis de componentes principales (PCA) para modelar las repuestas impulsivas relativas a la cabeza (HRIRs). Varios autores, han utilizado diferentes tipos de preprocesamientos de la HRIRs antes de aplicar PCA y son disímiles los criterios para determinar la cantidad de componentes principales necesarias. En este trabajo se demuestra que el PCA aplicado a los valores complejos de la transformada de Fourier (HRTF) de las HRIRs es más efectivo que el PCA aplicado a la magnitud y al logaritmo de la magnitud de las HRTFs. Para componentes principales (PC) superiores a la séptima, la velocidad de decrecimiento del error medio cuadrático global (mean-square error: MSE) es mayor para el preprocesamiento propuesto. El estudio revela que las magnitudes espectrales de las 4 primeras PCs, muestran la influencia de la oreja en las HRTFs y que a partir de la 7 PCs se manifiesta la preponderancia del torso y de los hombros.

1. INTRODUCCIÓN

Las respuestas impulsivas (HRIRs) entre diferentes posiciones de una fuente sonora y ambos tímpanos de un oyente, reflejan el efecto de filtraje causado por las estructuras anatómicas de cabeza, torso y orejas. En un estudio pionero, Wightman and Kistler (1989) comprobaron que los sujetos podían localizar fuentes sonoras virtuales usando auriculares, con la misma precisión con que localizaban fuentes sonoras reales en campo libre. El procedimiento consistió en filtrar una señal anecoica, con las mencionadas respuestas impulsivas. Con anterioridad, [Mehrgardt y Mellert \(1977\)](#) habían demostrado que las HRIRs son secuencias de fase-mínima y que el resto de fase, esto es, la diferencia entre la fase total de las HRIRs medidas menos la fase-mínima, es casi lineal con la frecuencia e igual a un simple retardo. Estas evidencias permitieron desarrollar un modelo simplificado conocido como: fase-mínima-más-retardo (minimum-phase-plus-delay) ([Kulkarni et al., 1999](#)). La respuesta impulsiva de fase-mínima se puede obtener apelando a la propiedad que relaciona el logaritmo de la magnitud con la fase ([Oppenheim et al., 1989](#)). Más adelante, [Kistler y Wightman \(1992\)](#) propusieron un modelo basado en el Análisis de Componentes Principales (PCA) y en las respuesta impulsivas de fase-mínima. El procedimiento consistió en aplicar el PCA al logaritmo de las magnitudes de las HRTFs de un grupo de sujetos. El PCA descompone el espectro de las HRTFs en un conjunto de funciones básicas o componentes principales (PCs). Estas PCs pueden ser consideradas formas espectrales básicas del espectro de las HRTFs medidas de forma tal, que pueden ser reconstruidas con la suma pesada de estas PCs. Los pesos definen la contribución relativa de cada PC al espectro de las HRTFs. El propósito del trabajo mencionado fue el de determinar cuantas PC son necesarias para reconstruir las HRTFs de un sujeto del grupo, sin que se degraden aquellas características espectrales fundamentales que un individuo utiliza para localizar una fuente sonora con precisión. Desafortunadamente en este trabajo, los autores no expresan porque utilizaron como datos de entradas al PCA el logaritmo de la magnitudes de las HRTFs. En un trabajo reciente ([Hugeng et al., 2010](#)) se realizó un estudio exhaustivo sobre el efecto que tienen sobre los resultados del PCA, los distintos tipos de preprocesamiento de las HRIRs. Ellos concluyeron que el preprocesamientos más efectivo en el dominio de la frecuencia fue el de las magnitudes de las HRTFs. Sin embargo, ellos no analizan las HRTFs complejas (magnitud y fase). En este trabajo, se estudia la influencia de tres tipos de preprocesamiento antes de aplicar el PCA: el logaritmo de las magnitudes de las HRTFs, propuesto en el trabajo pionero de [Kistler y Wightman \(1992\)](#); el más eficaz según [Hugeng et al. \(2010\)](#), esto es, el de las magnitudes de las HRTFs y el propuesto aquí: las HRTFs complejas. Se analizan las funciones básicas o componentes principales (PCs) obtenidas de los valores complejos y se determina cual de los tipos de preprocesamientos analizados es el más eficaz, esto es, cual requiere de menos PCs para reproducir las características relevantes de las HRTFs.

2. ANÁLISIS DE COMPONENTES PRINCIPALES

Para este estudio se utilizó la base de datos de HRIRs del CIPIC ([Algazi et al., 2001a](#)). Consta de HRIRs medidas a 35 sujetos a las entradas de ambos conductos auditivos bloqueados. Para cada sujeto, se realizaron mediciones en 1250 posiciones de la fuente sonora. La ubicación de la fuente se especifica por el ángulo de acimut φ y el ángulo de elevación θ referidos a un eje que pasa por ambos oídos denominado eje interaural. Son secuencias de 200 puntos muestreadas a 44100 Hz.

En primera instancia, se obtuvieron las HRTFs de los 35 individuos, mediante las transformadas de Fourier de las HRIRs. Luego, se armaron tres grupos de 87500 HRTFs cada uno (1250

posicionesx35sujetosx2oídos) con los diferentes tipos de procesamiento: con las HRTFs complejas (ComHRTF); con los módulos de las HRTFs (MagHRTF) y con los logaritmos base 10 de los módulos de las HRTFs (LogHRTF). Finalmente, se aplicó el PCA a cada tipo de preprocesamiento y se obtuvieron para cada uno de ellos las PCs, que son funciones de la frecuencia, y los pesos W que dependen del sujeto s , del acimut φ y de la elevación θ de la fuente sonora. Una HRTFs de cualquiera de los tipos, se puede reconstruir totalmente calculando:

$$H(f, \varphi, \theta, s) = H_0(f) + \sum_{i=1}^N PC_i(f) \cdot W_i(\varphi, \theta, s) \quad (1)$$

en donde H_0 es la media global del conjunto; N es la totalidad de las PC; f la frecuencia y s uno de los sujeto en particular. Para conseguir una efectiva reducción de la dimensión de las HRTFs medidas, se debe determinar un número $L \ll N$, de manera que las HRTFs reconstruídas conserven aquellas características espectrales que, desde el punto de vista perceptual, son relevantes (Ramos et al., 2007) (Ramos y Tommasini, 2008) (Kistler y Wightman, 1992).

Se denominará HRTFmed, HRTFcom, HRTFmag y HRTFlog a las HRTFs medidas y a las reconstruidas a partir de ComPCA, MagPCA y LogPCA respectivamente. Es oportuno advertir, que las HRTFs reconstruidas con la Ecuación (1), tienen la misma dimensión de los datos de entrada al PCA. Para evaluar objetivamente el rendimiento de los tres tipos de preprocesamiento, se utilizó el error medio cuadrático global (mean-square error: MSE), el cual es definido como (Chen et al., 1995) (Hugeng et al., 2010):

$$MSE(\%) = 100 \cdot \frac{\|h_{med} - h_{pca}\|^2}{h_{med}} \quad (2)$$

donde h_{med} son las HRTFs medidas y h_{pca} son las reconstruidas por PCA. Como las HRTF-com son complejas, se utilizó el módulo para calcular el MSE. En la Fig. 1 se grafica el MSE en función de la cantidad de PCs empleadas en la reconstrucción y para los tres tipos de preprocesamiento.

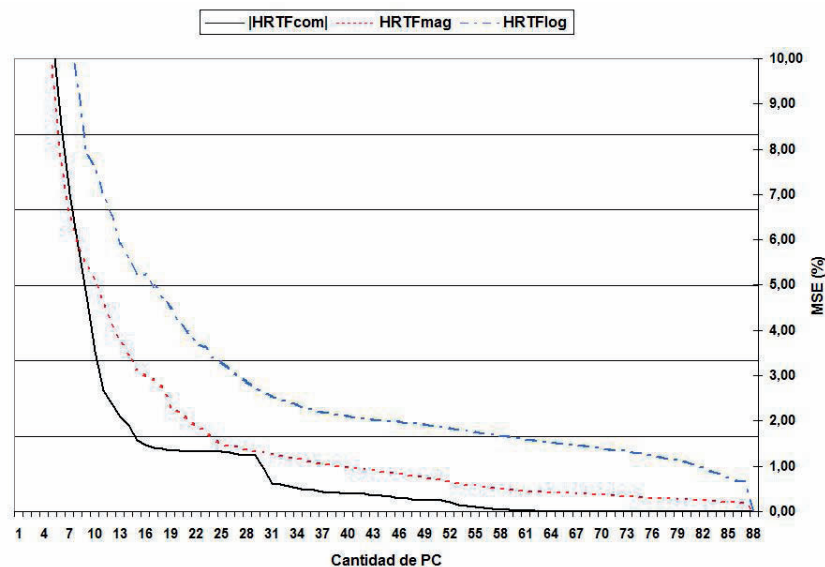


Figura 1: MSE global para todas las posiciones, sujetos y oídos obtenidos de los diferentes tipos de preprocesamientos

Claramente se observa que la velocidad de decrecimiento del MSE del HRTFcom es mayor a las de las otras dos y que a partir de las 7 PCs el error es menor. Para 12 PCs las diferencias de los errores son máximas. En la Tabla 1, se muestran los MSE cuando las HRTFs se reconstruyen con 12 PCs:

Tipo	HRTFcom	HRTFmag	HRTFlog
MSE(%)	2.37	4.15	6.55

Tabla 1: MSE global para 12 PCs

3. ANÁLISIS DE LAS PCS DE LAS HRTFS COMPLEJAS

Shaw (1997) identificó seis resonancias características originadas en las orejas y las llamó modos normales. De mediciones realizadas a 10 sujetos, determinó los valores promedios de las frecuencias de resonancias, los niveles de presión sonora y direcciones de incidencias (Fig. 2).

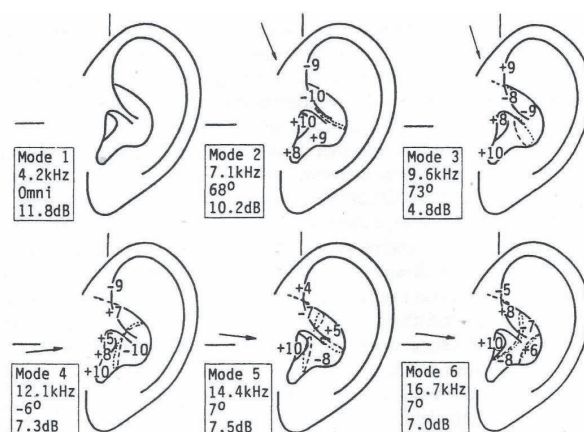


Figura 2: Modos normales (resonancias) medidos por Shaw (1997)

En las magnitudes de los espectros de las cuatro primeras PCs, PC#1-PC#4 (Fig. 3), varias resonancias son notables. Algunas de las resonancias observadas en esta figura, coinciden con aquellos modos normales.

PC#1 exhibe un máximo entre los 4 y 5 kHz compatible con el modo 1. Según Shaw (1997), este modo es independiente de la dirección de incidencia de la onda sonora, esto es, omnidireccional. PC#2 muestra tres picos de la misma amplitud en 3.8 kHz, 7.6 kHz y 12.7 kHz aproximadamente. Los dos últimos coinciden con los modos 2 y 4 (7.1 kHz y 12.1 kHz para incidencias de 68° y -6° respectivamente). El modo 3 es claro en PC#3 a 9.6 kHz (9.6 kHz a 73° de incidencia). Un pico en 14.1 kHz compatible con el modo 5 es visible en PC#4 (14.4 kHz a 7° de incidencia). El modo 6 (16.7 kHz a 7°) no es observado ya que esta fuera del rango de frecuencias analizadas. Más adelante, se demostrará que estas PCs, también cumplen con los ángulos de incidencias determinados por Shaw (1997). Podemos adelantar, sin embargo, que las cuatro primeras PCs reflejan, al menos, la influencia de las orejas en las HRTFs.

Los perfiles espectrales de los módulos de las PCs cambian a partir de PC#5, mostrando picos y mínimos (notches) en apariencia periódicos. Se determinaron las frecuencias medias de los picos y mínimos de las PC#5 a la PC#12. Si las frecuencias promedios de los picos y mínimos de una PC en particular son comparables, se podrá afirmar que el espectro es más o menos periódico. En la Tabla 2 se muestran las frecuencias promedios de ambos:

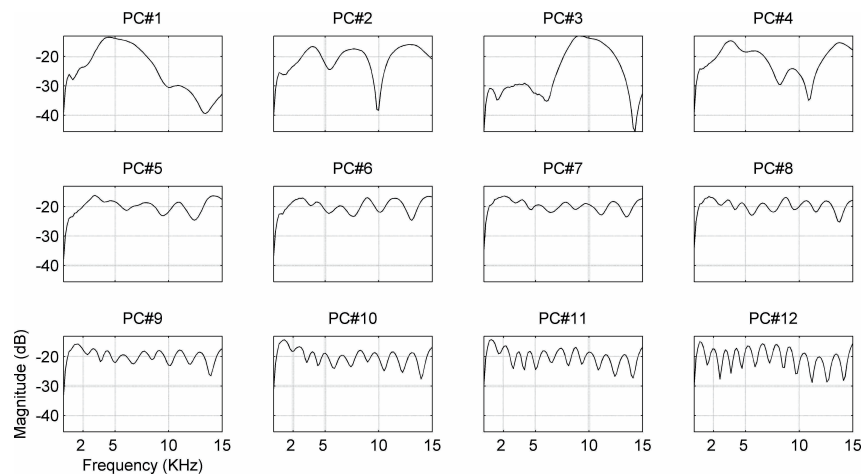


Figura 3: Magnitud de las 12 PCs más significativas obtenidas de las HRTFs complejas

PC#	f_{peak}	f_{notch}
5	2756.3	2813.7
6	2153.3	2368.7
7	2032.7	2067.2
8	1837.5	1837.5
9	1624.2	1648.8
10	1485.8	1507.3
11	1378.1	1397.3
12	1274.8	1274.8

Tabla 2: Frecuencias promedios de los picos y mínimos

A medida que el número de la PC aumenta, se asemejan las frecuencias de los picos y de los notch. La diferencia entre la f_{peak} y f_{notch} para la PC#5 es igual a 57.5 Hz (2.0 %); para la PC#6 es de 215.33 (10 %) y a partir de PC#7, las diferencias entre las f_{peak} y f_{notch} es menor a 34.5 Hz (1.7 %), y es igual a cero para la PC#12. La disminución sistemática de estas diferencias ha partir de PC#7, nos permite inferir que la magnitud de los espectros de estas PCs son periódicos, similares a la respuesta de un típico filtro "comb". Estos picos y notch periódicos son el resultado de la interferencia a la entrada de los conductos auditivos, de dos ondas coherentes que recorren caminos diferentes: la directa y de reflexiones producidas en diversas partes de cuerpo.

En la Fig. 4 se muestra una secuencia de imágenes de las HRTFs reconstruidas con una hasta con 12 PCs del oído derecho de un sujeto de la base y en el plano vertical (acimut igual a cero). El eje vertical representa la frecuencia; el eje horizontal la elevación de la fuente sonora en grados y la intensidad de los grises, la magnitud en dB. Las áreas brillantes se corresponden con valores de mayor presión sonora (resonancias) y las más oscuras a valores de menor presión (notches). Las elevaciones varían entre -45° y 230° en pasos de 5.6° . Una elevación de 0° significa que la fuente sonora está ubicada al frente del sujeto, en un plano horizontal que contiene a un eje imaginario que pasa por la entrada de ambos conductos auditivos (eje interaural). Elevaciones negativas, al frente y por debajo del plano mencionado; 90° justo encima de la cabeza del sujeto; 180° detrás del sujeto en el plano horizontal y posiciones mayores a esta, detrás y debajo de sujeto.

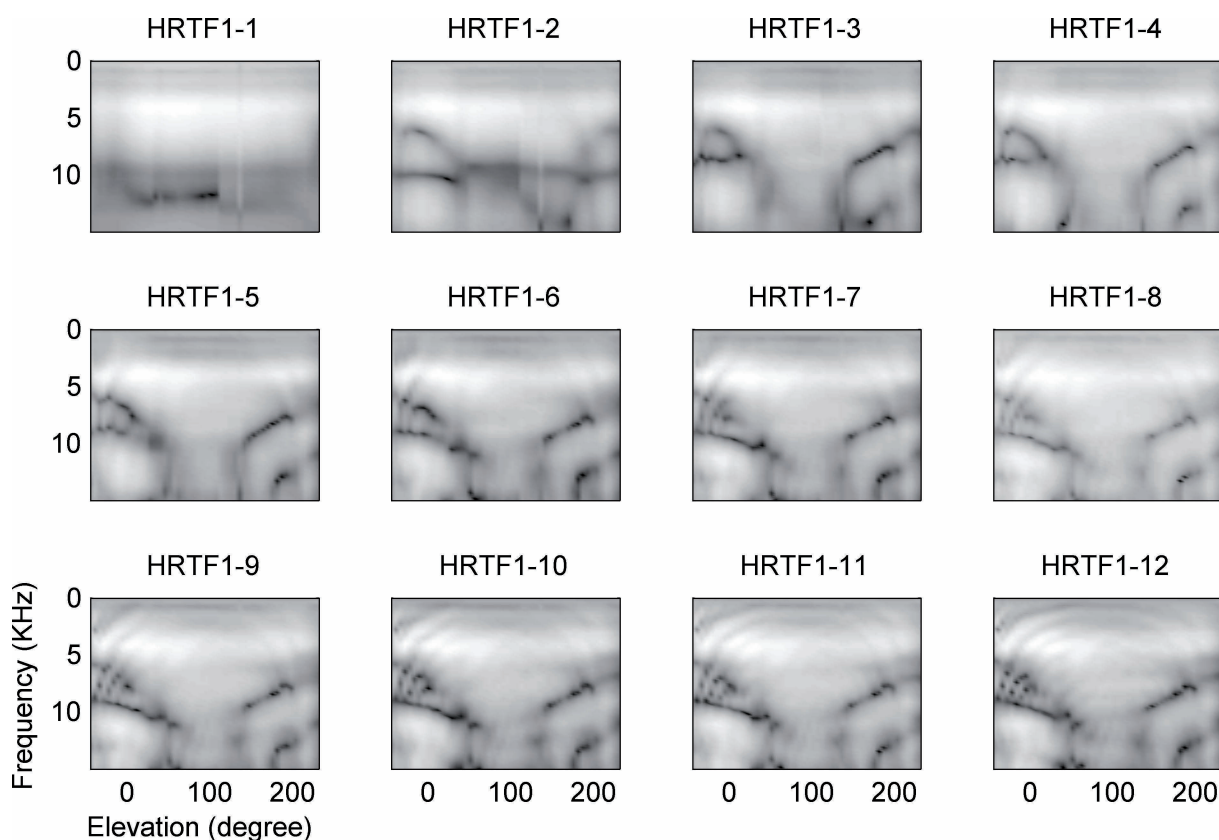


Figura 4: HRTFs reconstruída con una a 12 PCs

En HRTF1-1 (HRTFs reconstruidas con una PC), alrededor de los 4 kHz, es notable una amplia área brillante que es independiente de la elevación. Como ya se mencionó, esta resonancia la aporta PC#1 (modo 1). En la HRTF1-2 emerge otra resonancia aproximadamente a 0° para frecuencias mayores a 10 kHz. Esta es una de las resonancia que aporta PC#2 (modo 4 a 12.1 kHz a -6°). Otra de las resonancias que aporta PC#2 es la del modo 2 (7.1 kHz a $+68^\circ$) que es menos notorio en la figura. PC#4 aporta a la HRTF1-4, otro de los modos horizontales: el 5 (14.4 kHz at $+7^\circ$). La interacción de este modo con el modo 4, provoca una zona más brillante entre los 11.5 kHz y 13.5 kHz en 0° aproximadamente. Entre los 6 kHz y los 10 kHz y entre -45° y $+45^\circ$ se observa un mínimo que varía sistemáticamente con la elevación. Este “notch” se origina por difracción/reflexión en la pared posterior de la concha de las orejas (Shaw et al., 1968) (Lopez-Poveda y Meddis, 1996). El desplazamiento sistemático de la frecuencia central, tiene un rol preponderante en la localización de fuentes sonoras en el plano vertical (Iida et al., 2007). En las sucesivas HRTFs y por debajo de los 3 kHz, son notables unas formas de arco dependientes de la elevación de la fuente, que se completan a medida que el número de PC aumenta. Los arcos son aproximadamente simétricos respecto a $+80^\circ$ y el filtro comb de más baja frecuencia, se produce en $+90^\circ$. Algazi et al. (2001b) determinó que estas formas observadas en las HRTFs por debajo de los 3 kHz, se deben a la interacción del sonido directo y las reflexiones producidas en el torso y hombros. Demostró también, que estos patrones observados para frecuencias menores a 3 KHz proporcionan claves adicionales o secundarias para localizar una fuente en el plano vertical. Por lo expuesto, las 12 PCs más significativas, tienen en cuenta el comportamiento de las HRTFs tanto en alta como en medias y bajas frecuencias.

4. COMPARACIÓN DE LOS TRES TIPOS DE PREPROCESAMIENTOS

Se mostrará con un ejemplo ilustrativo como el preprocesamiento de las HRTFs complejas es más eficiente, requiriendo menos PCs para preservar más características de las HRTFs medidas. Se comparó $20\log_{10}(|HRTF_{med}|)$ de dos sujetos de la base de datos, con $20\log_{10}(|HRTF_{com}|)$, $20\log_{10}(HRTF_{mag})$ y $20(HRTF_{log})$, reconstruidas con 12 PCs (Fig. 5). Corresponden al oído derecho para posiciones de la fuente en el plano vertical, esto es, acimut igual a cero. En general, los valores máximos de presión sonora coinciden, pero hay marcadas diferencias en los valores mínimos (ver la barra de grises a la derecha de cada gráfico). Las $HRTF_{mag}$ y $HRTF_{log}$ muestran escasa actividad o nula por debajo de 3 kHz y los notchs de las $HRTF_{log}$ en las zonas de actividad de las orejas (>4 Kz) son de mayor ancho de banda. Ahora, las $HRTF_{com}$ muestran una marcada similitud con las HRTFs medidas, mostrando detalles que en las dos anteriores no se observan.

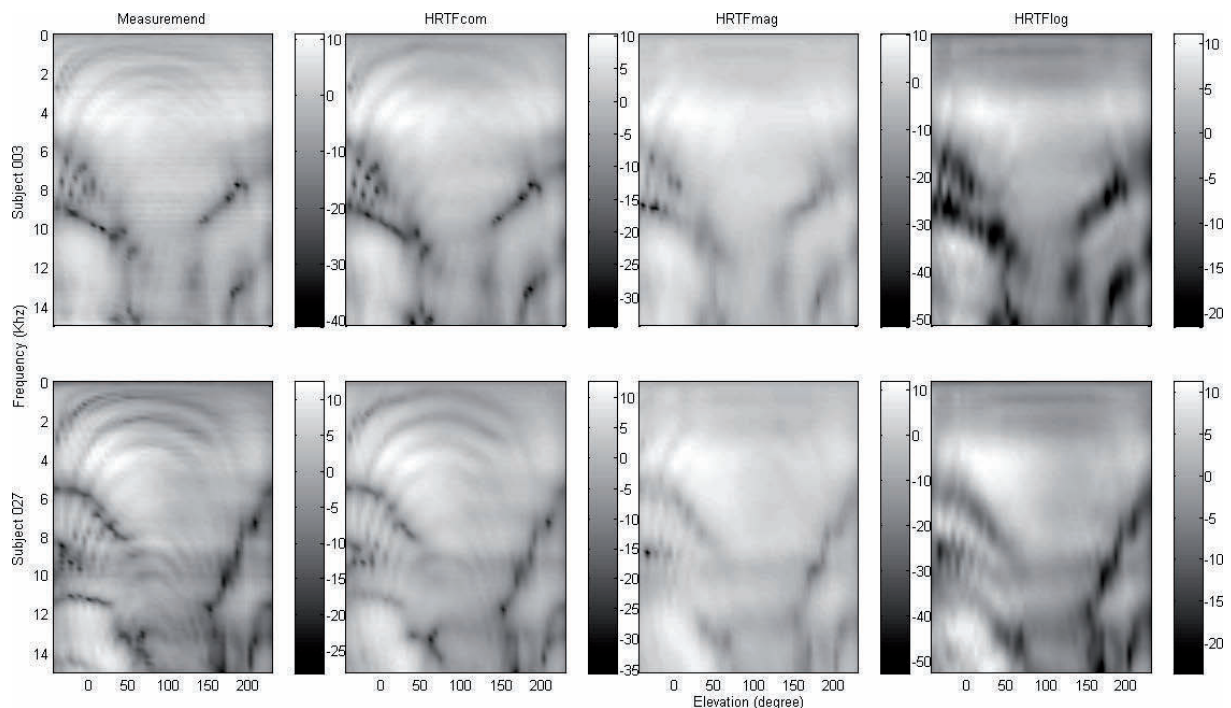


Figura 5: Ejemplo ilustrativo de la efectividad del modelo propuesto

En la Tabla 3 se muestra los MSE para el oído derecho de 5 sujetos de la base de datos en el plano vertical, para los tres tipos de preprocesamiento.

Subject	HRTFcom	HRTFmag	HRTFlog
003	5.03	6.17	14.78
010	4.41	6.18	23.15
018	5.49	9.54	15.81
020	5.24	7.05	15.28
027	6.24	8.68	23.36

Tabla 3: MSE promedio en el plano vertical (%)

La tendencia de los MSE en el plano vertical es igual que los MSE globales respecto a todas las direcciones y sujetos, mostrados en la Tabla 1.

5. CONCLUSIONES

Se ha demostrado que el PCA aplicado a las HRTFs complejas de un grupo de sujetos y para todas las posiciones alrededor de la esfera, es más eficaz que el PCA aplicado a la magnitud y al logaritmo de la magnitud de las HRTFs. La velocidad de decrecimiento del MSE del primer método es mayor a partir de la séptima PC. Las HRTFs del plano vertical reconstruidas con las cuatro primeras PCs, muestran, para frecuencias superiores a 4 Khz, la influencia de la orejas en las HRTFs. De igual manera, la reconstrucción de las HRTFs con la PC#5 hasta la doce, exhiben el comportamiento de las HRTFs en frecuencias menores a 3 Khz debido a la influencia del torso y hombro. De la comparación de las HRTFs obtenidas en el plano vertical con los tres tipos de preprocesamiento, mostró que con 12 PCs, los MSE de las HRTFcom para varios sujetos de la muestra son significativamente menores y la similitud con las HRTFs medidas es notable.

REFERENCIAS

- Algazi V., Duda R., Thompson D., y Avendano C. The CIPIC HRTF database. páginas 99 –102. 2001a. doi:10.1109/ASPAA.2001.969552.
- Algazi V.R., Avendano C., y Duda R.O. Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109(3):1110–1122, 2001b. doi:10.1121/1.1349185.
- Chen J., Veen B.D.V., y Hecox K.E. A spatial feature extraction and regularization model for the head-related transfer function. *The Journal of the Acoustical Society of America*, 97(1):439–452, 1995. doi:10.1121/1.413110.
- Hugeng, Wahab W., y Gunawan D. Effective preprocessing in modeling head-related impulse responses based on principal components - google académico. *Signal Processing: An International Journal (SPIJ)*, 4(4):201–212, 2010.
- Iida K., Itoh M., Itagaki A., y Morimoto M. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, 68(8):835–850, 2007. ISSN 0003-682X. doi:10.1016/j.apacoust.2006.07.016.
- 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. ISSN 0001-4966. PMID: 1564200.
- Kulkarni A., Isabelle S.K., y Colburn H.S. Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105(5):2821–2840, 1999.
- Lopez-Poveda E.A. y Meddis R. A physical model of sound diffraction and reflections in the human concha. *The Journal of the Acoustical Society of America*, 100(5):3248–3259, 1996. doi:10.1121/1.417208.
- 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.
- Oppenheim A.V., Schafer R.W., Buck J.R., et al. *Discrete-time signal processing*, volumen 2. Prentice hall Upper Saddle River^ eN. NJ, 1989.
- Ramos O., Calvo G., y Tommasini F. Modelo acústico de cabeza y torso mediante análisis de componentes principales. *Mecánica Computacional*, 26:46–58, 2007.
- Ramos O.A. y Tommasini F.C. Simplificación de las funciones de transferencia de cabeza mediante análisis de componentes principales. *Mecánica Computacional*, 27:431–442, 2008.
- Shaw E.A.G. *Acoustical Features of the Human External Ear in Binaural and spatial hearing in real and virtual environments*. Lawrence Erlbaum, 1997.

Shaw E.A.G., Teranishi R., et al. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J. Acoust. Soc. Am*, 44(1):240–249, 1968.