

## SIMPLIFICACIÓN DE LAS FUNCIONES DE TRANSFERENCIA DE CABEZA MEDIANTE ANÁLISIS DE COMPONENTES PRINCIPALES

Oscar A. Ramos<sup>a,b</sup> y Fabián C. Tommasini<sup>a,b</sup>

<sup>a</sup> *Centro de Investigación y Transferencia en Acústica, UTN, Regional Córdoba-Unidad Asociada del CONICET. Maestro Lopez esq. Cruz Roja Argentina, Córdoba, Argentina.*

[oramos@scdt.frc.utn.edu.ar](mailto:oramos@scdt.frc.utn.edu.ar)

<sup>b</sup> *Miembros del CONICET*

**Palabras Clave:** funciones de transferencia de cabeza, componentes principales.

**Resumen.** Una onda sonora antes de alcanzar los tímpanos de un oyente, sufre una serie de transformaciones al interactuar con el torso, los hombros y especialmente con la cabeza y los pabellones auriculares. Estas transformaciones son diferentes para cada oído y dependen del ángulo de incidencia de la onda. Ha sido demostrado que aspectos bien definidos de las ondas sonoras que llegan a los tímpanos, llevan información que un ser humano utiliza para localizar con precisión una fuente sonora en el espacio. Las Funciones de Transferencia (HRTF) medidas entre una fuente sonora y los tímpanos contienen dichas transformaciones.

Es un trabajo previo, se desarrolló un modelo de estas transformaciones mediante el método de Análisis de Componentes Principales (PCA), a partir de HRTF complejas medidas a 5 sujetos en la Universidad de Wisconsin por el Dr. Fred Wightman. Pruebas psicoacústicas preliminares revelaron que la utilización de entre 20 y 25 funciones básicas son suficientes para reconstruir una HRTF que no puede ser distinguida de la original.

Es propósito del presente trabajo, avanzar en la interpretación físico-acústicas y sus implicancias perceptuales de las Componentes Principales con la finalidad de diseñar estrategias que reduzcan aún más la cantidad de componentes necesarias.

## 1 INTRODUCCION

Las ondas sonoras antes de alcanzar los tímpanos de un oyente son alteradas por el torso, los hombros y especialmente por la cabeza y los pabellones auriculares (orejas). Según el ángulo de incidencia las transformaciones son diferentes para cada oído. Las respuestas impulsivas (Head Related Impulse Response: HRIR) medidas en campo libre entre una fuente sonora situada en diferentes posiciones, contienen dichas transformaciones.

Ha sido un interés permanente de los investigadores, estudiar la naturaleza de tales transformaciones y su implicancia para la audición binaural. Se ha demostrado que las ondas sonoras que llegan a los tímpanos, transportan información que un ser humano utiliza para localizar con precisión una fuente sonora en el espacio (Blauert, 1997; Wightman and Kistler, 1989a, b).

La diferencia de tiempo (ITD) y la diferencia de nivel (ILD) entre ambos oídos, son determinantes para localizar una fuente sonora. Estas dos claves binaurales son conocidas como claves *primarias*. Sin embargo, existen lugares del espacio en que tanto las ITD e ILD son constantes, dando origen a *zonas de confusión* en donde al ser humano le es imposible establecer con precisión la ubicación de la fuente (Blauert, 1997). En estos casos, juegan un rol preponderante el pabellón auricular, los hombros y el torso que transforman el espectro de las ondas sonoras, según la dirección de incidencia. Este efecto de filtrado natural se debe principalmente a las múltiples resonancias originadas en las protuberancias de las orejas y a las reflexiones en el torso y en los hombros, provocando cancelaciones y refuerzos en lugares bien definidos del espectro. Estas claves *secundarias* son denominadas claves monoaurales y son importantes para determinar la elevación de una fuente sonora por arriba y debajo de un plano horizontal que pasa por ambos oídos. Se habrá podido advertir que, individuos diferentes producirán transformaciones diferentes de las claves mencionadas. Por ejemplo, las diferencias interaurales de tiempo y nivel varían con el diámetro de la cabeza y los espectros con el tamaño y orientación de las orejas, entre otras razones.

Es de interés para la simulación acústica virtual, modelar dichas transformaciones de manera convincente, esto es, que un sujeto “inmerso” en un *entorno auditivo virtual*, experimente la vivencia auditiva que percibiría en un entorno real. En este contexto, dos son los problemas a resolver. Por un lado, las diferencias inter-individuos y por el otro la eficiencia del modelo en relación a los recursos computacionales necesarios.

## 2 ANTECEDENTES

Desde fines de los 80 hasta nuestros días, la implementación *eficiente*<sup>1</sup> de modelos del oyente acaparó el interés de numerosos investigadores. Algazi et al. (2001b, 2002), emplean modelos teóricos tanto de la cabeza como de las orejas y modelan estos elementos con filtros paramétricos lineales. Han validado los filtros implementados con mediciones de las ondas sonoras en cabezas y orejas artificiales. Otros autores, utilizan filtros de distintos tipos y orden (FIR y IIR) que son ajustados a mediciones de las HRTF (Kulkarni, and Colburn, 2004; Ramos, et. al., 2006).

Modelar la dependencia espacial de las HRTF o HRIR mediante un conjunto de *funciones básicas* ha sido intentado por numerosos autores. Martens (1987) aplicó el método de Análisis de Componentes Principales (PCA) a HRTF preprocesadas con filtros de banda crítica, pero sólo en el plano horizontal. Kistler and Wightman (1992) aplicaron el mismo procedimiento a los logaritmos de las magnitudes de las HRTF. Demostraron que con la combinación lineal de

---

<sup>1</sup> En el sentido en que los tiempos de cálculo y los requerimientos de memoria sean los mínimos

cinco funciones básicas se podía reconstruir el logaritmo de las magnitudes y la fase con un retardo lineal. Chen et al. (1995) utilizaron la expansión de Karhunen–Loève para obtener funciones básicas ortogonales a partir de los valores complejos (magnitud y fase) de HRTF medidas en una cabeza artificial y un gato. Los estudios previos se realizaron en el dominio de la frecuencia. Estudios similares se realizaron en el dominio del tiempo. Wu et al. (1997) emplearon la expansión de Karhunen–Loève a HRIR obtenidas de un solo sujeto. Más recientemente, Park y Hwang (2008) aplican PCA a HRIR de un extenso grupo de sujetos solamente en el plano medio. Determinaron que 12 componentes son suficientes para reconstruir las HRIR, basándose en el error porcentual, fijado arbitrariamente en 5 %, entre las HRIR medidas y las reconstruidas.

Por nuestra parte (Ramos et. al, 2007), aplicamos el método de PCA a un conjunto de HRTF medidas en 5 sujetos y realizamos pruebas psicoacústicas de discriminación entre estímulos sintetizados con las HRTF originales y las generadas por el modelo. El estudio reveló que la utilización de pocas funciones básicas –entre 20 y 25- son suficientes para preservar las características fundamentales de fase y magnitud de las HRTF medidas, de manera tal que los sujetos no pudieron discriminar entre los estímulos generados con las originales y las reconstruidas. Se concluyó que la diferencia al cuadrado entre las magnitudes de las originales y las reconstruidas, integrada a lo largo de todo el espectro, debía resultar menor a 2 dB. Decíamos además que, la cantidad de funciones básicas estaba determinada por el ajuste del oído oculto por la cabeza y que se debían realizar estudios complementarios para reducir aún más el orden del modelo.

Es propósito del presente trabajo, avanzar en la interpretación físico-acústica y sus implicancias perceptuales de las Componentes Principales con la finalidad de diseñar estrategias que reduzcan ese orden. Se utilizaron el mismo grupo de HRIR medidas a 5 sujetos en la Universidad de Wisconsin por el Dr. Fred Wightman.

El trabajo está organizado de la manera siguiente. En el punto 3, se introduce al lector en las bases del Análisis de Componentes Principales y su aplicación a nuestro caso; en el punto 4 se analiza el error de ajuste y en el punto 5, se discute el significado de las componentes principales.

### 3 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El análisis de Componentes Principales es un método estadístico que pretende reducir la dimensionalidad de un conjunto de datos intercorrelacionados, resaltando sus características comunes. Cuando la dimensión de los datos es grande, el método permite además, develar patrones disimulados en el conjunto. Consiste en proyectar el conjunto de datos empíricos u observaciones sobre un conjunto de vectores ortonormales o *funciones básicas*, en el sentido en que la covarianza común es máxima.

El método permite obtener un conjunto de funciones básicas o componentes principales (CP) y un conjunto de *pesos* que reflejan la contribución relativa de cada CP para reconstruir un miembro en particular de los datos originales. Cada CP captura en mayor o menor grado algún aspecto común del conjunto.

Si  $H$  es la matriz de los datos empíricos de dimensión  $n \times m$ ;  $CP$  la matriz de las funciones básicas de dimensión  $m \times m$  y  $W$  la matriz de los pesos de dimensión  $n \times m$  tendremos:

$$H = CPW^t \quad (1)$$

es decir, que cualquier miembro de  $H$  se reconstruye a partir de la combinación lineal de  $CP$  y la transpuesta de  $W$ .

En nuestro caso, las HRIR fueron medidas en 505 posiciones alrededor de una esfera de 1,5 metros de diámetro con una resolución de 10 grados. Para definir la ubicación de la fuente se utiliza un sistema de coordenadas esféricas. El origen es la intersección de tres planos: el plano horizontal o acimutal, que contiene una línea imaginaria que une ambos oídos (eje binaural); el plano medio o vertical, que divide la cabeza simétricamente en derecha e izquierda y el plano frontal, perpendicular a los dos anteriores. Cada posición está definida por los ángulos de acimut  $\varphi$  y elevación  $\Phi$ . Si  $\varphi=0$  y  $\Phi=0$ , la fuente está ubicada al frente del sujeto en el plano acimutal y el vertical. Si  $\varphi$  es negativo la fuente está a la izquierda del plano vertical y si  $\varphi$  es positivo a la derecha. De igual manera, si  $\Phi$  es negativo la fuente está por debajo del plano horizontal y si  $\Phi$  es positivo por arriba. A partir de ahora definiremos la posición de la fuente por el par (acimut, elevación), esto es, por ejemplo, (-90,0).

Las elevaciones varían entre -50 grados (debajo del plano horizontal que pasa por ambos oídos) y +90 grados (encima de la cabeza) mientras que los acimuts varían entre 0 y -180 grados (a la izquierda del plano medio) y entre 10 y 170 grados (a la derecha del plano medio). Cada HRIR son secuencias de 256 puntos que muestreadas a 44100 Hz, dan una extensión de 5,8 milisegundos.

Antes de aplicar el procedimiento, las HRIR fueron convertidas al dominio de la frecuencia mediante la transformada de Fourier (HRTF). La resolución fue de 172,4 hz y el ancho de banda fue limitado a 15000 hz. Por lo tanto, se utilizaron las primeras 88 componentes en frecuencias. En la tabla 1 se muestra la estructura de los datos:

Sujeto	Posición	Frecuencias			
AFW	1	f1	f2	...	fm
	...	...	..	...	...
	p	f1	f2	...	fm
SJX	1	f1	f2	...	fm
	...	...	..	...	...
	p	f1	f2	...	fm
SOS	1	f1	f2	...	fm
	...	...	..	...	...
	p	f1	f2	...	fm
SOU	1	f1	f2	...	fm
	...	...	..	...	...
	p	f1	f2	...	fm
SOW	1	f1	f2	...	fm
	...	...	..	...	...
	p	f1	f2	...	fm

Tabla 1: Estructura de los datos empíricos

Cada sujeto contribuye con 505 (posiciones) x 2 (oídos)= 1010 espectros complejos de 88 componentes de frecuencias. Por lo tanto, la matriz global con los datos de los 5 sujetos será:  $5 \times 1010 = 5050$  filas ( $n$ ) por 88 columnas ( $m$ ) igual a 444.400 números complejos.

Se utilizó la función *princomp* de Matlab y se obtuvo la matriz de CP de 88 componentes principales de 88 frecuencias y la matriz  $W$  de 88 pesos (uno por cada CP) por 5050 posiciones ( $n$ ). Cualquier HRTF del conjunto, se reconstruye totalmente haciendo:

$$\hat{H}(f, \varphi, \theta) = H_0(f) + \sum_{i=1}^{88} CP_i(f) W_i(\varphi, \theta) \quad (2)$$

en donde  $H_0$  es la media global del conjunto que se fue calculada y restada de cada componente antes de aplicar el procedimiento.  $H_0$  refleja los rasgos comunes e independientes de la posición de la fuente sonora.

Para conseguir una efectiva reducción de la dimensión de los datos originales, se debe determinar un número  $L$ , bastante menor a 88, que permita reconstruir una HRTF dada. Diferentes criterios se pueden adoptar para establecer  $L$ . Algunos autores lo hacen fijando un valor para el error rms entre las HRTF medidas y las aproximadas. En nuestro caso el valor de  $L$  fue establecido como resultado de pruebas psicoacústicas de discriminación de sonidos procesados con las HRTF originales y las aproximadas por el modelo (Ramos et. al, 2007).

#### 4 ANÁLISIS DEL ERROR DE APROXIMACIÓN

Como se vio, el plano medio divide la cabeza del sujeto en dos hemisferios simétricos. Al oído que está en el mismo hemisferio que la fuente se lo denomina ipsilateral, mientras que al oído opuesto: contralateral.

##### 4.1 Distribución del Error en función de la posición de la fuente

Para cada posición de la fuente y para cada oído, se calculó el error medio cuadrático de las diferencias entre las magnitudes de las HRTF medidas y las aproximadas.

$$Error_n = \sqrt{\frac{1}{m} \sum_{m=1}^{m=88} [20 \log |H_n(f_m)| - 20 \log |\hat{H}_n(f_m)|]^2} \quad (3)$$

Luego, se computó la media aritmética de los errores para cada posición y de todos los sujetos. En la Figura 1 se muestran los resultados para el oído derecho y el izquierdo. En primer lugar, se observa que sus perfiles son similares, pero complementarios. El error es menor para el oído ipsilateral (el oído que está en el mismo hemisferio que la fuente) y mayor para el oído contralateral (el oído que está en el hemisferio opuesto a la fuente y oculto por la cabeza). Esto es, los errores del oído derecho son mayores para acimuts negativos y menores para acimuts positivos (Figura 1a). Lo contrario ocurre para el oído izquierdo (Figura 1b). En segundo lugar, se ve que los errores son mayores para elevaciones negativas (debajo del plano horizontal) y que disminuyen sistemáticamente para elevaciones positivas, siendo mínimo cuando la fuente esta sobre la cabeza (elevación igual a 90 grados). A modo de síntesis, el error máximo se produce por debajo del plano horizontal y para cuando el oído está oculto por la cabeza.

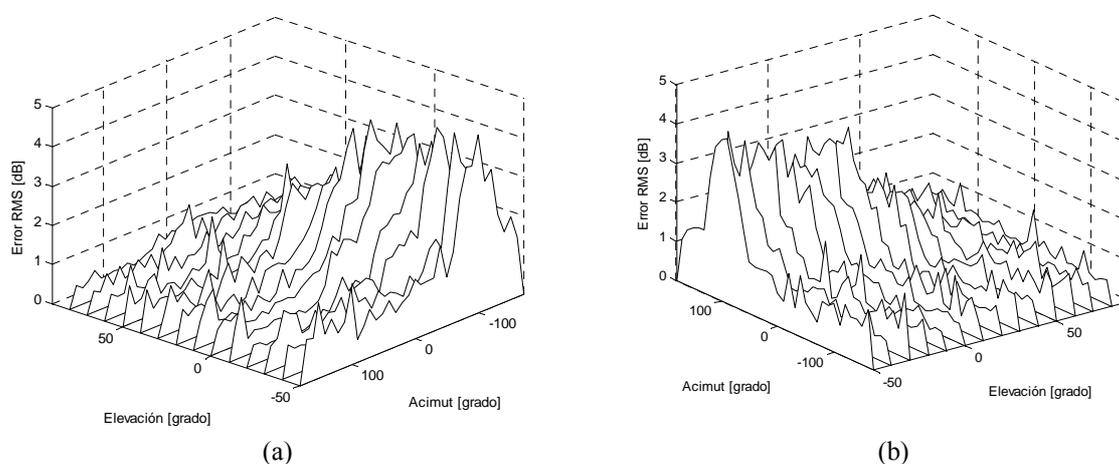


Figura 1: Distribución del error medio de magnitud entre los 5 sujetos en función de la posición de la fuente sonora: a) oído derecho y b) oído izquierdo

Hemos comprobado que la distribución del error de las magnitudes es característica e independiente de la cantidad  $L$  de CP utilizadas para reconstruir las HRTF. Lo mismo ocurre con el error de fase, razón por la cual no se presentan aquí. Se observó también, que a medida que se aumenta el número de CP, los errores máximos se producen en el oído contralateral cuando la fuente está directamente enfrentada al oído ipsilateral en un rango que varía en  $\pm 20$  grados y para valores de elevación que varían entre  $-50$  y  $+10$  grados.

Note que, los errores máximos se producen cuando el efecto de difracción o *sombra acústica* provocada por la cabeza es máximo y en consecuencia los perfiles espectrales de las HRTF son complicados y difíciles de ajustar con pocas CP.

La cuestión es: ¿cuál es el error aceptable desde la perspectiva psicoacústica? Como se vio más arriba, en estudios previos que realizamos (Ramos et. al, 2007), se arribó a que el error debía estar por debajo de 2 los dB. De los gráficos se desprende que los errores medios superan los 2 dB para elevaciones comprendidas entre  $-50$  y  $+10$  y acimuts entre  $70$  y  $100$  grados para el oído izquierdo. Para el oído derecho ocurre algo similar con errores comprendidos entre 2 y 3 dB por debajo del plano horizontal y acimuts próximo a los  $+90$  grados.

Esta asimetría entre los errores del oído ipsilateral y el contralateral es una constante en todos los modelos de las HRTF obtenidos por otros autores e independiente del método utilizado (Kulkarni and Colburd, 2004).

#### 4.2 Errores en el plano horizontal y vertical en función de la frecuencia

En la Figura 2 se grafican los errores medios de las magnitudes de los cinco sujetos en función de la frecuencia en el plano horizontal y el vertical. El oído mostrado es el derecho y las CP utilizadas son 25. En el plano horizontal (Figura 2a) los errores mayores están concentrados en la región de los acimuts negativos y para frecuencias superiores a los 6000 Hz. Esto es, para posiciones de la fuente sonora en el hemisferio opuesto al oído analizado y en la zona del espectro en que la difracción provocada por la cabeza es importante. El color preponderante en esta zona es el azul claro y corresponde a errores comprendidos entre 1 y 2 dB.

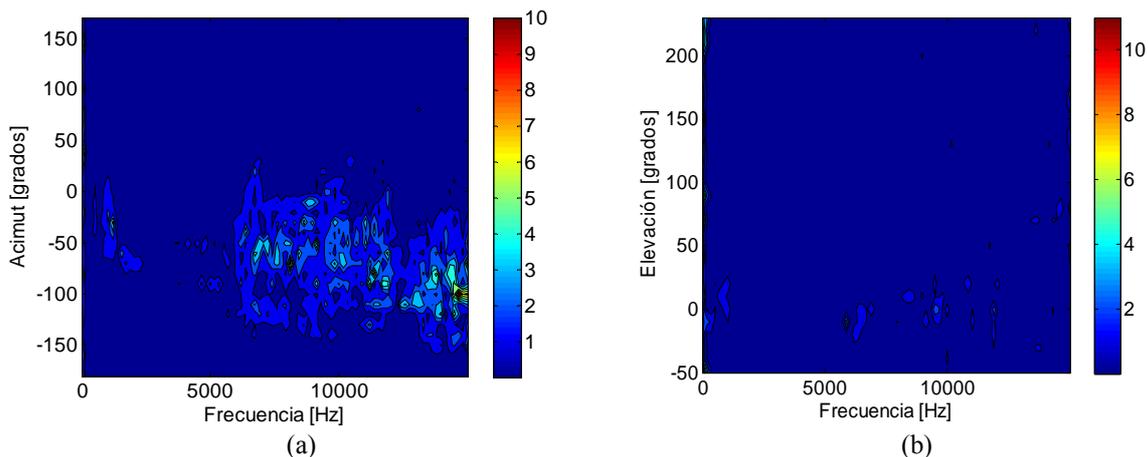


Figura 2: Distribución del error medio de magnitud entre los 5 sujetos en función de la frecuencia: (a) en el plano horizontal y (b) en el plano vertical

Los errores mayores (zonas verdes, amarillas y rojas) no presentan ningún patrón o tendencia. El error máximo (8,23 dB) se da para un acimut de  $-100$  grados y 14643 Hz.

El error es menor a 1 dB (azul intenso) para acimut positivos (la fuente está directamente enfrentada al oído) y para acimut negativos (cuando la fuente está detrás) en todo el rango de frecuencias.

En el plano vertical (Figura 2b), en general, los errores son menores. En la región del azul intenso los errores son menores a 0.5 dB. El error mayor de casi 4 dB, se produce para una elevación de -10 grados a una frecuencia de 5857 Hz. Es difícil establecer un patrón de los errores, sin embargo se puede notar que son mayores para elevaciones entre  $\pm 20$  grados para frecuencias comprendidas entre 6 y 12000 Hz (para esas elevaciones, las reflexiones en el torso son importantes).

En ambos planos e independiente de la posición, los errores son menores para frecuencias comprendidas entre 1000 Hz y 6000 Hz.

Los errores de fase, no mostrados aquí, están distribuidos más uniformemente que los de magnitud. Los errores mínimos son del orden de la milésima de radianes alcanzado un valor máximo de casi 2 radianes en zonas muy puntuales.

### 4.3 Errores en función del número de CP y posición de la fuente

Como una manera de estimar la variación del error en función de la cantidad de CP utilizadas (orden del modelo), se calculó el error promedio para los cinco sujetos del oído contralateral y del ipsilateral. Para estimar el error del *oído contralateral* se agrupó el error del oído derecho para ubicaciones de la fuente en el hemisferio izquierdo con el error del oído izquierdo para la fuente ubicada en el hemisferio derecho. Para evaluar *el oído ipsilateral*, se agrupó los errores de cada oído para la fuente ubicada en el mismo hemisferio. En la Figura 3 se muestran los resultados para la magnitud y la fase.

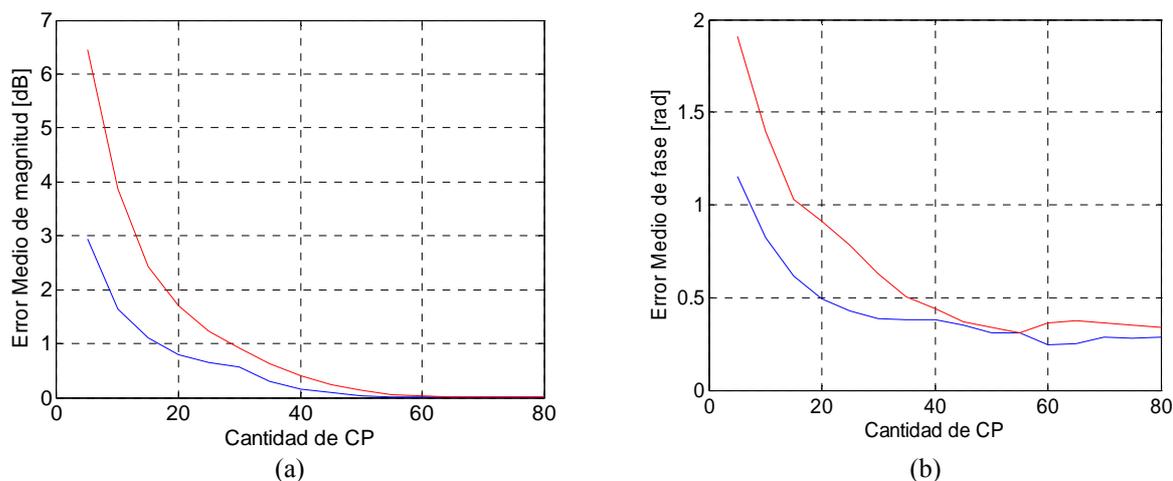


Figura 3: Error medio global de a) magnitud y b) fase del oído contralateral (azul) y del oído ipsilateral (rojo) en función de la cantidad de CP utilizadas en la reconstrucción

En ambos gráficos, el error del oído contralateral (rojo) es siempre mayor que el del oído ipsilateral (azul). Después de 60 CP, los errores de magnitud (Fig. 3a) se estabilizan y se igualan; mientras que los errores de fase (Fig. 3b) parecen estabilizarse después de 70 CP. El error del oído contralateral, decrece más rápido entre la 5 y 15 CP, mientras que el oído ipsilateral lo hace entre la 5 y la 10. El error del oído ipsilateral toma valores menores a 2 dB para 10 CP, mientras que los errores contralaterales recién alcanzan estos valores para 20 CP. Esta diferencia en la convergencia del error, gobierna el orden definitivo del modelo. Estos resultados objetivos del error concuerda con los obtenidos de las pruebas psicoacústicas realizadas en un estudio anterior (Ramos et. al, 2007) en que la utilización de 20 a 25 CP eran suficientes y que el error de ajuste era menor a 2 dB.

Del análisis previo, surge la necesidad de indagar como se comporta el error en regiones particulares.

Cuando la fuente sonora está situada en el plano vertical (Fig. 4a), los oídos nunca quedan ocultos por la cabeza y tanto las diferencias interaurales de tiempo como las de nivel son iguales a cero. Ambas curvas se comportan de igual manera y el error medio de la magnitud es menor a 2 dB para 10 CP.

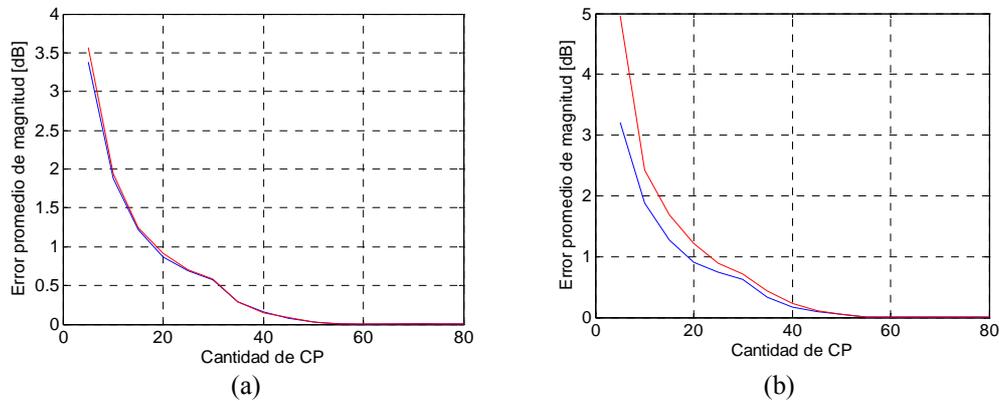


Figura 4: Error medio en el (a) plano vertical y (b) plano igual ITD, del oído ipsilateral (azul) y del oído contralateral (rojo) en función de la cantidad de CP

A medida que la fuente sale del plano vertical, el efecto de difracción provocado por la cabeza, aumenta progresivamente en el oído oculto siendo máxima cuando la fuente está enfrentada al otro oído. Por ejemplo, en la Figura 4b se muestra el error cuando la fuente se ubica en un círculo que resulta de la intersección de la esfera con un plano paralelo al vertical. Este círculo tiene la particularidad de que las diferencias de tiempo interaurales son constantes. Al ser el ITD constante, esta zona es potencialmente productora de confusión por la irrelevancia de esta clave primaria. En la Figura 4b se ve que el error producido en el oído contralateral aumenta respecto al del plano vertical y es menor a 2 dB recién para 15 CP. Obsérvese que el error en el oído ipsilateral es igual al de la Figura 4a, esto es, es menor a 2 dB para 10 CP.

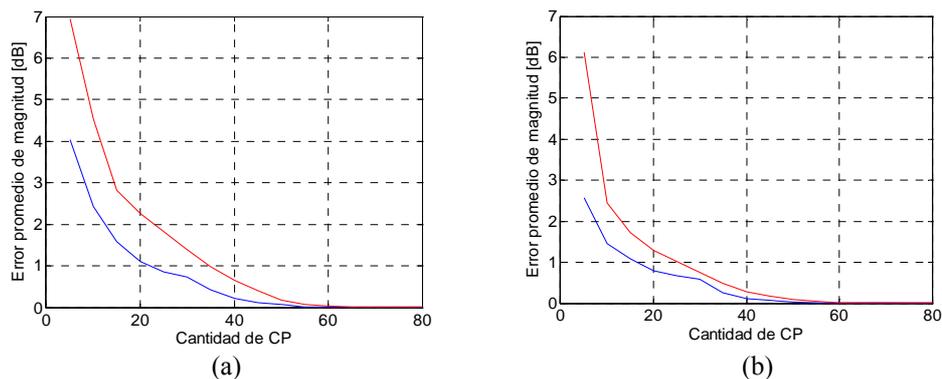


Figura 5: Error medio del oído ipsilateral (azul) y del oído contralateral (rojo) en función de la cantidad de CP: a) elevación -50 grados y b) elevación +50 grados.

En la Figura 5 se grafican los errores para ubicaciones de la fuente en sendos planos paralelos situados por debajo (elevación -50 grados) y por arriba del plano acimutal (+50 grados). El error del oído ipsilateral aumenta levemente para la elevación de -50 grados (Figura 5a) alcanzando el valor de 2 dB entre 10 y 15 CP. Sin embargo el deterioro del oído contralateral es notable necesitando casi 25 CP para alcanzar los 2 dB. Por el contrario, para elevaciones por encima de la línea binaural (Figura 5b) los errores en ambos oídos, disminuyen significativamente.

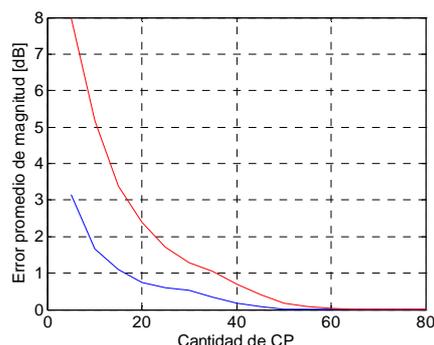


Figura 6: Error medio en el plano horizontal del oído ipsilateral (azul) y del oído contralateral (rojo) en función de la cantidad de CP

Finalmente, el error en el plano acimutal es mostrado en la Figura 6. El error en el oído contralateral es el mayor a los ya vistos, mientras que por el contrario el error del ipsilateral alcanza el valor de 2 dB para ocho CP. Obsérvese además, que la separación entre ambas curvas es mayor, resultando mayor la diferencia de CP necesarias para reconstruir ambos oídos.

En conclusión: para las regiones analizadas, el error del oído ipsilateral es menor a 2 dB si se utiliza sólo 10 CP, salvo para posiciones por debajo del plano acimutal en donde las reflexiones en el torso y la difracción producidas en los hombros son importantes. Por el contrario el error del oído contralateral está gobernado por el fenómeno de difracción producido por la cabeza, siendo mayor para ubicaciones de la fuente en el plano acimutal y para cuando la fuente enfrenta directamente al oído ipsilateral. La diferencia de orden o cantidad de CP requeridas es mayor también en este plano.

## 5 ANÁLISIS DE LAS CP

Para completar el estudio, analizaremos las Componentes Principales. Reacuérdesse que las CP son valores complejos en función de la frecuencia. En la tabla 2 se muestra la contribución, en porcentaje, que hace cada CP al conjunto. La columna *parcial* es el acumulado de las varianzas de las CP anteriores. Las primeras cinco CP representan el 91,41 % de la varianza total del conjunto de datos originales. A partir de la CP9 la contribución individual de cada CP es menor del 1% alcanzando con la 16 una varianza acumulada del 99,52 %. Sin embargo, a pesar de lo elevado de este valor, se demostró en el punto anterior que para que las HRTF reconstruidas no puedan ser discriminadas de las originales era necesario emplear entre 20 y 25 CP en algunas regiones.

En la figura 7a se grafican las magnitudes de las primeras cinco CP (91,41 %). Se observa, en general, la existencia de picos y ceros marcados. En primer lugar y en las cinco, se ven picos que varían entre los 2400 Hz y los 2700 Hz y corresponden a la principal resonancia del conducto auditivo. En la CP2 existe otro pico alrededor de los 8 kHz; en la CP3 en 10 kHz y en la CP4 en 12 kHz. Estas resonancias coinciden aproximadamente con los modos de resonancias de la oreja definidos por Shaw como modos 2, 3, 4 (7.1 kHz; 9.6 kHz y 12.1 kHz). Además define el modo 1 en 4.1 kHz y los modos 4 y 5 en 14.4 kHz y en 16.7 kHz respectivamente. Los modos 1 y 5 no se observan mientras que el modo 5 está fuera de nuestro rango de análisis. Shaw determina estas resonancias a partir de mediciones promediadas en oídos de humanos y de modelos mecánicos del pabellón auricular (Shaw, 1997).

Ceros muy notorios se notan en 10.1 kHz en CP2 y en 9.3 kHz en CP4 debido a cancelaciones producidas en las protuberancias de las orejas también.

CP	Varianza	Acumulada	CP	Varianza	Acumulada
1	48,438	48,44	9	0,773	97,48
2	19,506	67,94	10	0,538	98,02
3	12,365	80,31	11	0,438	98,46
4	7,002	87,31	12	0,326	98,79
5	4,099	91,41	13	0,256	99,04
6	2,669	94,08	14	0,203	99,24
7	1,577	95,66	15	0,154	99,40
8	1,053	96,71	16	0,120	99,52

Tabla 2: Contribución de cada CP a la varianza total

En la Figura 7b se grafican las correspondientes CP en el dominio del tiempo, obtenidas por la transformada inversa de Fourier de las CP de la Figura 7a. Se observa que la mayor energía está concentrada en los primeros 500 microsegundos que se corresponde con reflexiones y resonancias compatibles con la dimensión de las protuberancias de la oreja. (Algazi et. al, 2001b).

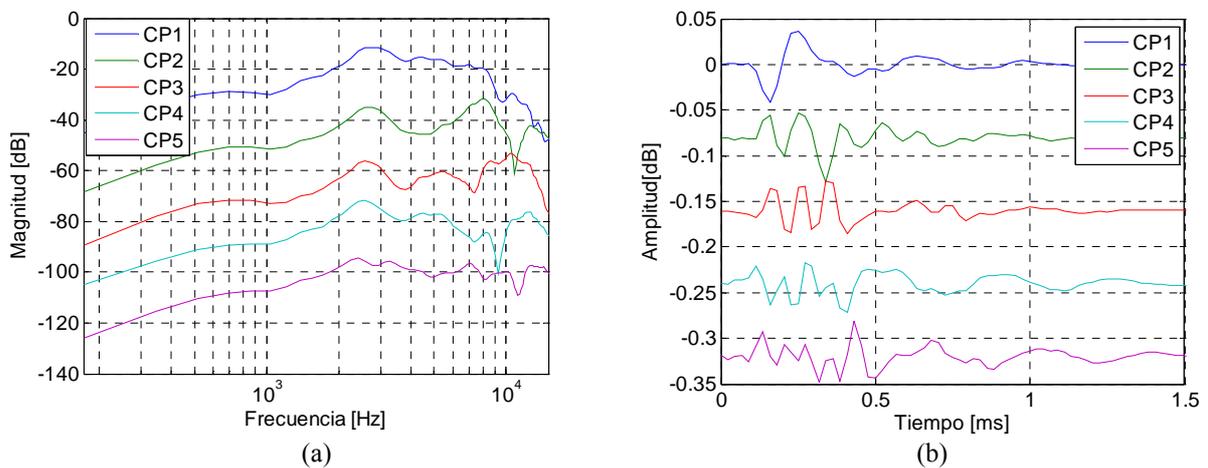


Figura 7: Magnitud (a) y respuesta temporal (b) de CP1 a la CP5

A partir de CP5 (no mostradas aquí) existe una zona de transición en donde conviven la influencia de la oreja y las reflexiones en los hombros y son difíciles de aislar e interpretar. Sin embargo a partir de la CP10 la influencia de las reflexiones en el torso son notables. En la Figura 8 se muestran desde la CP10 hasta la CP15 tanto en el dominio de la frecuencia como en el del tiempo. Se nota claramente la existencia de reflexiones casi discretas que se desplazan en el tiempo hasta 1 ms en CP15 (Figura 8b). Esta evidencia es confirmada por el perfil de las magnitudes en la Figura 8a que se asemejan a una respuesta típica de filtros “peines” (comb filter). El retardo entre el sonido directo (CP1) y la reflexión mostrada en CP15 es alrededor de 900 microsegundos correspondiéndose con 31 cm compatible con reflexiones en el torso.

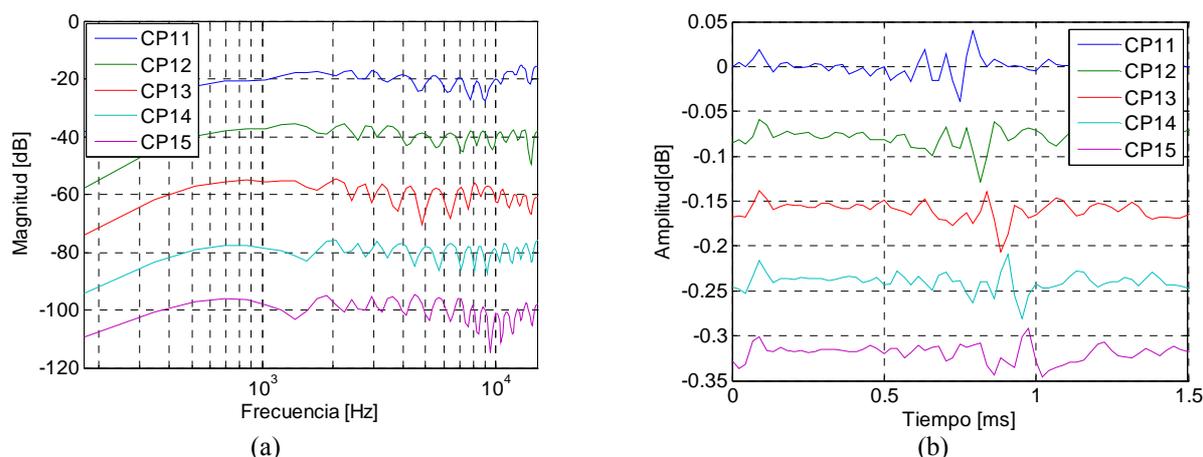


Figura 8: Magnitud (a) y respuesta temporal (b) de CP11 a la CP15

## 6 ANÁLISIS DE RESULTADOS Y CONCLUSIONES

Con la finalidad de obtener un modelo eficiente del receptor humano, se aplicó el método de Análisis de Componente Principales a un conjunto de HRIR medidas a cinco sujetos. Estudios psicoacústicos previos determinaron que la utilización de entre 20-25 CP eran suficientes para conservar las características de magnitud y fase de las HRIR originales. Para reducir aún más el orden del modelo se intenta, con el estudio presente, develar el comportamiento del error de ajuste con la posición de la fuente sonora y el significado físico-acústico de las CP. A modo de conclusión diremos:

- Sólo 10 CP son necesarias y suficientes para mantener el error del oído ipsilateral por debajo de 2 dB independiente de la posición de la fuente.
- El error del oído contralateral es altamente dependiente de la posición de la fuente, siendo mínimo en el plano vertical y máximo en el plano acimutal y por debajo de dicho plano. Además, el error disminuye sistemáticamente desde el plano acimutal hacia elevaciones positivas.
- La distribución del error con respecto a la frecuencia del oído contralateral y en el plano acimutal crece a partir de los 6000 Hz.
- En las 5 primeras CP es notoria la resonancia en el conducto auditivo que es independiente de la posición de la fuente. Ellas también reflejan los modos de resonancias originadas en las protuberancias y cavidades de los pabellones auriculares (Shaw, 1997).
- A partir de la CP10 la influencia de las reflexiones en el torso son notorias.

La transformada de Fourier analiza una señal con filtros de ancho de banda constante a lo largo de todo el rango de frecuencias. El comportamiento del aparato auditivo humano es algo diferente. Experimentos psicoacústicos demostraron que el análisis en frecuencia que lleva a cabo el oído es equivalente a filtros de porcentaje constante (banda-crítica), esto es, que la relación entre la frecuencia central y el ancho de banda de los filtros es una constante (Zwicker, 1961). En baja frecuencia el ancho de banda es pequeño y crece a medida que la frecuencia aumenta. En consecuencia, es de esperar que los finos detalles del espectro que resultan de aplicar la FFT a las HRIR, sean suavizados si son preprocesadas con filtros que simulen el análisis en frecuencia que realiza el sistema auditivo (especialmente en alta frecuencias).

Ha sido demostrado que la respuesta del conducto auditivo es independiente de la posición de la fuente (Mehrgardt and Mellert, 1977). Para eliminar de las HRTF la respuesta del conducto auditivo se debe normalizar cada HRTF con la respuesta conocida como *campo difuso*. Esta se obtiene calculando la media cuadrática de la totalidad de las HRTF.

Ambos procedimientos tienen por finalidad eliminar de las HRTF aspectos que son independientes de la posición de la fuente y perceptualmente irrelevantes. De esta manera, la variaciones presentes en las HRTF serán suavizadas y se podrá lograr un mejor ajuste con menos CP.

## 7 REFERENCIAS

- Algazi, V., Duda, R., Thompson, D. and Morrison, R. Structural Composition and decomposition of HRTF, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY, 103–106, 2001b.
- Algazi, V., Duda, R., Duraiswami, R., Gumerov, N. and Tang, Z. Approximating the head-related transfer function using simple geometric models of the head and torso,” *J. Acoust. Soc. Am.* 112, 2053–2064, 2002.
- Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA. The MIT Press, 1997.
- Chen, J., Van Veen, B. D. and Hecox, K. E. A spatial feature extraction and regularization model for the head-related transfer function. *J. Acoust. Soc. Am.* 97 439–452, 1995.
- Hwang, S. and Park, Y. Interpretations on principal components analysis of head-related impulse responses in the median plane, *J. Acoust. Soc. Am. Express Letters*, 123, (4), E165-E171, 2008.
- Kistler, D. and Wightman, F. A model of head-related transfer functions based on principal components analysis minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91, 3, 1637-1647, 1992.
- Kulkarni, A. and Colburn, S. Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am.*, 115(4), 1714-1728, 2004.
- Martens, W. L. Principal components analysis and resynthesis of spectral cues to perceived direction. *Proceedings of International Computer Music Conference*, San Francisco, CA, 1987
- Mehrgardt, S. and Mellert, V. Transformation characteristics the external human ear, *J. Acoust. Soc. Am.* 61, 1567–1576, 1977.
- Ramos, O., Jasá, V., Gorgas, J. and Céspedes Daza, D. Simulación Binaural de Espacios Acústico. *Memorias del V Congreso Iberoamericano de Acústica*. Santiago de Chile. Chile, 2006.
- Ramos, O., Calvo, G. and Tommasini, F. Modelo Acústico de Cabeza y Torso Mediante Análisis de Componentes Principales. *Mecánica Computacional*, Vol. XXVI, pp 46-58, 2007.
- Shaw, E. Acoustical Features of the Human External Ear in *Binaural and Spatial Hearing in Real and Virtual Environments*. Editor’s Gilkey, R. and Andersosn, T. Lawrence Erlbaum Associates, Inc, 1997.
- Wightman, F. and Kistler, D. Headphone simulation of free-field listening I: Stimulus synthesis, *J. Acoust. Soc. Am.* 85, 858-867, 1989a.
- Wightman, F. and Kistler, D. Headphone simulation of free-field listening II: Psychophysical validation, *J. Acoust. Soc. Am.* 85, 868–878, 1989b.
- Wu, Z., Chan, F. H.Y., Lam, F. K., and Chan, J. C. K. A time domain binaural model based on spatial feature extraction for the head-related transfer function, *J. Acoust. Soc. Am.* 102, 2211–2218, 1997.
- Zwicker, E. Subdivision of the audible frequency range into critical bands, *J. Acoust. Soc. Am.* 33, 248.