Asociación Argentina
de Mecánica Computacional

Mecánica Computacional Vol XXXII, págs. 3005-3012 (artículo completo)
Carlos G. García Garino, Aníbal E. Mirasso, Mario A. Storti, Miguel E. Tornello (Eds.)
Mendoza, Argentina, 19-22 Noviembre 2013

PREPROCESAMIENTO EFICAZ PARA MODELAR LAS RESPUESTAS IMPULSIVAS DE CABEZA, MEDIANTE ANÁLISIS DE COMPONENTES PRINCIPALES: VALIDACIÓN SUBJETIVA

Oscar A. Ramos, Fabián C. Tommasini, Sebastián P. Ferreyra, Agustín Cravero y Martín Guido

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, psicoacústica**.**

Resumen. Es muy frecuente, tanto en el dominio de tiempo como en el de la frecuencia, el uso del análisis de componentes principales (PCA) para modelar las repuestas impulsivas relativas a la cabeza (HRIR). Varios autores, han utilizado diferentes tipos de preprocesamientos o formatos de las HRIR 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 de las HRIR (HRTF) es más efectivo que el PCA aplicado a la magnitud y al logaritmo de la magnitud de las HRTFs. Esta afirmación es validada mediante pruebas psico-acústicas de discriminación entre las HRTF medidas en el plano vertical y los tres tipos de preprocesamientos mencionados.

1 INTRODUCCION

Numerosos autores han usado el PCA para modelar y personalizar las funciones de transferencia de cabeza. Algunos aplican el PCA a las HRIR mientras que otros lo hacen a sus transformadas de Fourier (HRTF). Las HRIR/HRTF reconstruidas con un número dado de componentes principales (PC), difieren de las HRIR/HRTF medidas según el tipo de preprocesamiento o formato adoptado. Usualmente, el formato más utilizado es el logaritmos de las magnitudes de las HRTF (por ejemplo, Kistler y Wightman, 1992; Hu et al, 2007; Xu et al, 2009). Según Leung y Carlile (2009), el formato más óptimo (mayor varianza acumulada con menos PC) es el de las magnitudes lineales de las HRTF[1]. Ellos aportaron además que con 6 o 7 PC la varianza acumulada es mayor al 90%, y que son necesarias entre 10 a 20 PC o más, para que una persona logre localizar una fuente sonora en el espacio con gran precisión (Leung y Carlile, 2009; Breebaart, 2013). Hugeng et. al (2010), realizaron un exhaustivo estudio sobre las implicancias que el tipo de preprocesamiento de los datos tiene sobre los resultados obtenidos del PCA. Ellos demostraron también que el formato más optimo es el lineal y que el error medio cuadrático (MSE) global es menor. Últimamente, Hölzl (2012) comprobó que los resultados obtenidos por Leung y Carlile (2009) y Hugeng et al (2010) son independientes de la base de datos de las HRIR.

En este trabajo, se analiza un formato no estudiado hasta ahora: las magnitudes complejas de las HRTF y los resultados son comparados con el PCA aplicado a las magnitudes y a los logaritmos de las magnitudes de las HRTF. El estudio se realiza en el plano medio en donde el espectro de las HRTF es relevante. Se realizan pruebas psicofísicas de discriminación entre estímulos sonoros procesados con las HRIR medidas y con las derivadas de los tres formatos bajo estudio.

2 PCA

Para el estudio se utilizó la base de datos de HRIR del CIPIC (Algazi, et al, 2001). Consta de HRIR medidas a 35 sujetos a las entradas de sus conductos auditivos bloqueados y 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 inter-aural. Son secuencias de 200 puntos muestreadas a 44100 Hz.

En primera instancia se obtuvieron las HRTF de los 35 individuos, mediante la transformada de Fourier de sus HRIR. Luego, se armaron tres matrices con los siguientes preprocesamientos: con los valores complejos de las HRTFs; con las magnitudes de las HRTF y con el logaritmo de las magnitudes de las HRTFs y a cada una se le aplicó el PCA.

Se comprobó que las varianzas acumuladas de los tres formatos alcanzan casi el 90 % entre la sexta y séptima PCs. Los formatos complejo y lineal logran el 93 % en la octava PC mientras que el formato logarítmico obtiene ese valor de varianza recién en la doceava PCs. Estos resultados concuerdan con los reportados en artículos previos (Leung y Carlile, 2009; Hugeng et. al, 2010; Hölzl, 2012). Después de la octava PCs, la varianza acumulada del formato complejo, crece más rápidamente y la diferencia mayor con el formato lineal y el logarítmico, se alcanza entre la 12 y la 13 PC (en la doceava PC, la varianza para el formato complejo es 96.81 % versus el 95.68% para el formato lineal y el 93.49% para el formato logarítmico. Es oportuno advertir, que estas pequeñas diferencias pueden conducir a cambios significativos en la percepción de la ubicación de una fuente sonora en el espacio (Leung y Carlile, 2009).

[1] En ocasiones usaremos indistintamente el término "lineal", para distinguirlo del logarítmico

Para estimar a priori el nivel de ajuste entre las HRTF reconstruidas de los diferentes formatos y las HRTF medidas, se calculó el error medio cuadrático global (MSE) utilizado por otros autores en trabajos similares (ver Hugeng et al, 2010, por ejemplo). Se comprobó que el MSE del formato complejo, en menor en un 1.70 % y casi en un 2 % que el MSE del formato lineal y del logarítmico respectivamente.

En la Figura 1 se muestran para el plano vertical y en escala logarítmica, las HRTF medidas y las reconstruidas con 12 PC para los tres formatos y de dos sujetos de la base de datos tomados al azar. Puede ser comprobado, una notable similitud entre las HRTF medidas y las HRTF reconstruidas del formato complejo (HCom), mostrando detalles que no se observan en las HRTF reconstruidas a partir del formato lineal (HLin) y el formato logarítmico (HLog).

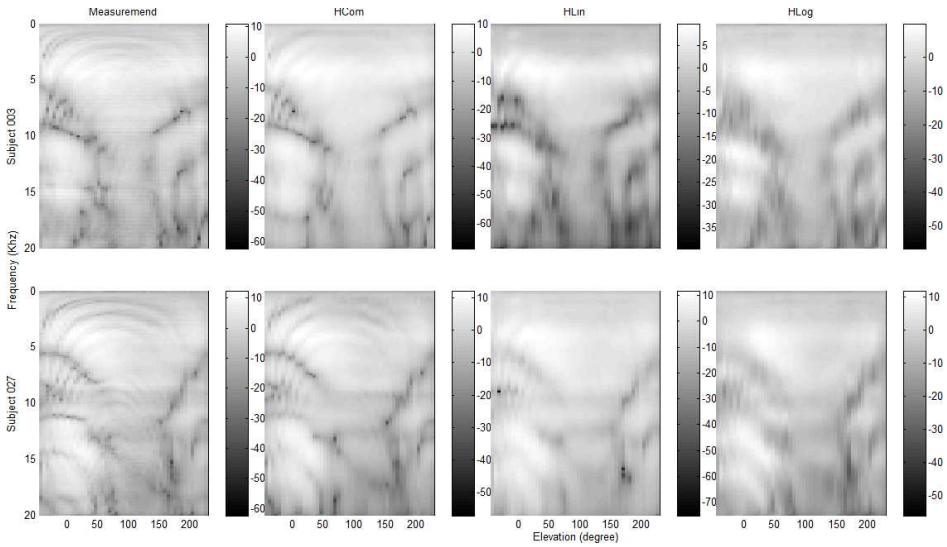

Figura 1: HRTFs medidas y reconstruidas con 12 PC en el plano medio para los diferentes formatos/preprocesamientos estudiados.

La pregunta es: estas pequeñas diferencias en las varianzas acumuladas y en el MSE son perceptualmente detectables?, teniendo en cuenta que algunos autores (por ejemplo: Scarpaci and Colburn, 2005) han demostrado la escasa correlación entre el MSE y el rendimiento psicofísico de las personas en experimentos de detección de fuentes sonoras?

3 OBTENCIÓN DE LAS HRIR DE FASE MÍNIMA

Mehrgardt and Mellert (1997) determinaron que las HRTF responden a un sistema de fase-mínima para las mayoría de las posiciones de la fuente sonora y que el resto de fase, esto es, la diferencia entre la fase total de las HRTF menos la fase-mínima, es casi lineal según la posición e independiente de la frecuencia. Por lo tanto, las HRTF pueden ser descompuestas en dos sistemas: un sistema de fase-mínima y un sistema "all-pass", en donde la magnitud de las HRTF es igual a la magnitud del sistema de fase-mínima (Oppenheim and Schafer, 1989). Estas evidencias permitieron desarrollar un modelo simplificado de las HRTF conocido como: fase-mínima-más-retardo (minimum-phase-plus-delay) (Kistler y Wightman, 1992; Kulkarni et al, 1999). En consecuencia, cada HRIR puede ser reemplazada por su respuesta asociada de fase-mínima más un retardo constante, que se corresponde con la diferencia de tiempo interaural (interaural time difference: ITD). Como nuestro estudio está limitado al plano medio, descartamos las ITD, que son cero o casi cero en este plano, y utilizamos las respuestas

impulsivas de fase-mínima solamente

Para obtener las respuestas impulsivas de fase-mínima asociadas a las HRTF reconstruidas con 12 PC de cada formato, se utilizó el ceptrum real:

$$\mathrm{hcom}(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(|\mathrm{HCom}|)\}\}.w(n)))\} \tag{1}$$

$$\mathrm{hlin}(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(\mathrm{HLin})\}\}.w(n)))\} \tag{2}$$

$$\mathrm{hlog}(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\mathrm{HLog})\}\}.w(n)))\} \tag{3}$$

y de las respuestas impulsivas de fase-mínima asociadas/derivadas de las HRIR medidas:

$$\mathrm{hmea}(n) = \mathrm{Re}\{\exp(F(\mathrm{Re}\{F^{-1}\{\log(|\mathrm{HRTF}|)\}\}.w(n)))\} \tag{4}$$

en donde Re es la parte real de un valor complejo; F y F^{-1} la transformada directa e inversa de Fourier respectivamente. Finalmente, w(n) es igual a 0 si n<0; igual a 1 si n=0 e igual a 2, si n>0 (Oppenheim and Schafer, 1989).

Para evaluar el grado de similitud entre las respuestas impulsivas de fase-mínima, asociadas a las HRIR medidas y las respuestas impulsivas de fase-mínima, obtenidas del PCA de los tres formatos, se calculó la función de correlación normalizada (Kulkarni et al, 1999):

$$\rho_{xy}(n) = \frac{\sum_{k=0}^{N} x(k)\,y(k+n)}{\sqrt{\sum_{k=0}^{N} x^2(k) \sum_{k=0}^{N} y^2(k)}} \tag{5}$$

y el índice de similitud o coherencia entre dos formas de onda definido como:

$$c = \max_{n} \left| \rho_{xy}(n) \right| \tag{6}$$

x(n) siempre fue la hmea(n) e y(n) fue hcom(n), hlin(n) e hlog(n). c es una estimación cuantitativa del grado de similitud o desviación entre x(n) e y(n). Si c=1, se dice que son coherentes o idénticas. Por el contrario, c expresa el grado de desviación entre x(n) e y(n) si es diferente a 1. En la Figura 2 se grafican los índices de coherencia promedio de las HRIR del plano medio de los 35 sujetos de la base de datos del CIPIC. Se advierte que, el índice promedio de similitud entre las hmea y las hcom es mayor en todo el plano medio.

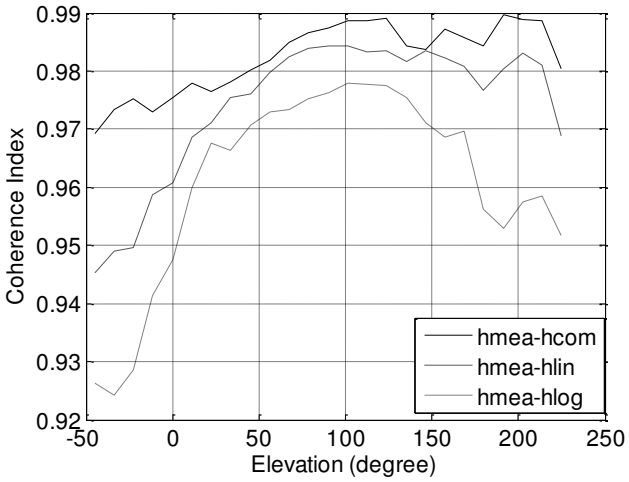

Figura 2: Índices de coherencias promedio entre las HRIR medidas de fase-mínima y las reconstruidas por PCA.

Nuevamente nos hacemos la misma pregunta: estas pequeñas diferencias a favor del formato complejo, son perceptualmente detectables y relevantes? Para responder esta pregunta se realizó el siguiente experimento.

4 EVALUACIÓN PERCEPTUAL

La prueba consistió en presentar a los participantes una secuencia de cuatro sonidos breves de 300 ms. de duración cada uno separados por un silencio de 300 ms. también. Tres de los estímulos sonoros se obtuvieron por convolución de un fragmento de ruido (ruido blanco con una rampa de crecimiento/decrecimiento cosenoidal al cuadrado de 50 ms) con las hmea y el cuarto estímulo por convolución con las respuestas impulsivas de fase-mínima de uno de los formatos estudiados. Este último ocupaba al azar el segundo o tercer intervalo. La tarea encomendada a los participantes fue que detectaran en cual de los dos intervalos, segundo o tercero, se encontraba el sonido diferente. Este paradigma de discriminación de elección forzada se conoce como de 4 intervalos y 2 alternativas: 4I-2AFC (Kulkarni et al, 1999; Kulkarni et al 2004) .

Participaron 10 sujetos voluntarios (5 varones y 5 mujeres) cuya edad estuvo comprendida entre los 19 y 29 años (promedio: 25 años). Todos los participantes no tenían experiencia previa en este tipo de experimentos. Para comprobar la condición audiológica de los participantes se les realizaron audiometrías de rango extendido hasta los 12 Khz, zona del espectro relevante en el plano medio.

Cada participante resolvió tres condiciones experimentales: hmea vs. hcom (COM); hmeas vs. hlin (LIN) y hmeas vs. hlog (LOG). Por razones de tiempo, el estudio se realizo en el plano medio para elevaciones comprendidas entre -45° y +90° en escalones de 10.25° (13 posiciones en total). Cada posición fue repetida 10 veces y fueron presentadas a los participantes al azar, resultando un total de 130 ensayos por condición experimental y por participante. Cada condición experimental duraba 15 minutos aproximadamente y el orden de administración de las tres condiciones experimentales fue al azar para cada participante. La HRIR utilizada en cada ensayo correspondía a un sujeto diferente de los 35 que componen la base de datos del CIPIC tomado al azar.

Los auriculares utilizados fueron los Sennheiser 570 de buena respuesta en frecuencia. Los estímulos fueron presentados a los escuchas mediante la E-MU's 0404 USB 2.0 Audio/MIDI

Interface de Creative Profesional.

4.1 Resultados y discusión

En la Figura 3a, se muestran el promedio de los porcentajes de las respuestas correctas de cada participante, a lo largo de todas las posiciones estudiadas del plano medio.

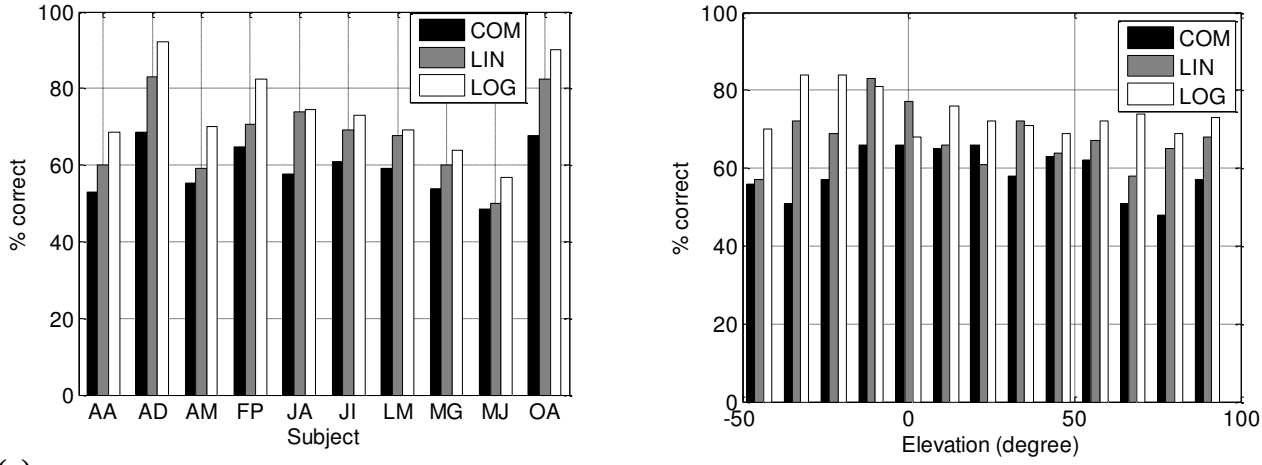

Figura 3: Promedio de las respuestas correctas de los 10 participantes: a) independiente de las posiciones y b) para cada posición.

A todos los participantes les fue más difícil resolver la condición COM que la condición LIN y la condición LOG (en ese orden). En otras palabras, tuvieron mayor dificultad para discriminar entre los sonidos procesados con las respuestas impulsivas de fase-mínima de las HRIR medidas y los sonidos procesados con las respuestas impulsivas de fase-mínima derivadas del formato complejo. Estos resultados acuerdan con los índices de coherencias calculados y graficados en la Figura 1. En la Figura 3b se ve el promedio de las respuestas correctas de todos los participantes para cada posición (10 repeticiones x 10 sujetos = 100 respuestas por posición). Se observa también acá que la tendencia es la misma que en el gráfico anterior: en promedio los participantes tuvieron más dificultad de discriminar en la condición COM que en las otras dos, en todas las posiciones evaluadas, excepto la de 23°.

Para determinar si estas diferencias son significativas se realizaron test de hipótesis t-Student que comparan dos medias de una sola cola. Los enunciados de la hipótesis nula (H_0) y alternativa (H_1) fueron:

$$H_0 : \mu_1 \geq \mu_2 \tag{7}$$

$$H_1 : \mu_1 < \mu_2 \tag{8}$$

en donde μ_1 y μ_2 fueron las medias de las respuestas correctas de todos los participantes para una posición en particular y para las condiciones experimentales comparadas. Se confrontaron las siguientes condiciones: LIN vs. LOG (LIN-LOG); COM vs. LIN (COM-LIN) y COM vs. LOG (COM-LOG) con un nivel de significancia del 0.05.

En la Tabla 1 se muestran los resultados de los t-test según ubicación de la fuente sonora. Las celdas con ceros marcan las posiciones de la fuente sonora en que se cumple la hipótesis nula H_0, es decir, que no hay diferencias significativas entre las condiciones experimentales comparadas. Por el contrario, las celdas con unos muestran las posiciones en que se cumple la hipótesis alternativa H_1. De la observación de la Tabla 1 se puede inferir que la condición LIN

es superior a la condición LOG en pocas posiciones de la fuente (23 %) y que la condición COM es superior a la condición LIN y LOG en un 30.8 % y 69.2 % de las posiciones respectivamente. En otras palabras: los estímulos sonoros procesados con las respuestas impulsivas de fase-mínima derivadas del formato complejo, son más difíciles de discriminar de los estímulos procesados con las HRIR medidas, que los estímulos procesados con las respuestas impulsivas de fase-mínima derivadas de la magnitud y del logaritmo de la magnitud.

Elevation (degree)	-45	-34	-23	-11	0	11	23	34	45	56	68	79	90	Average (%)
LIN-LOG	0	1	1	0	0	0	0	0	0	0	1	0	0	23,1
COM-LIN	0	1	0	1	0	0	0	1	0	0	0	1	0	30,8
COM-LOG	1	1	1	1	0	0	0	1	0	1	1	1	1	69,2

Tabla 1: Resultados de las pruebas de hipótesis. La celda con 0 significa que se cumple la hipótesis nula; por el contrario en las celdas con 1 la hipótesis nula se rechaza y es válida la hipótesis alternativa (ver texto)

5 CONCLUSIONES

Se ha demostrado que el PCA aplicado a los valores complejos de las HRTF, tiene algunas ventajas sobre el PCA aplicado a las magnitud y al logaritmo de la magnitud de las HRTF: a) la varianza acumulada crece más rápidamente a partir de la 8 PC y la diferencia con los formatos lineal y logarítmico es máxima entre la 12 y la 13 PC; b) el MSE global de las HRTF reconstruidas respecto a la HRTF medidas es menor para el formato complejo y c) el índice de coherencia entre las respuestas impulsivas de fase-mínima asociadas a las HRTF medidas y las deducidas de las HRTF reconstruidas a partir del formato complejo es mayor en todo el plano vertical. Se demostró, mediante pruebas psicoacústicas de discriminación, que las pequeñas diferencias numéricas de estos indicadores objetivos son perceptualmente detectables. Los participantes de las pruebas tuvieron una mayor dificultad para diferenciar los estímulos sonoros procesados con las HRTF reconstruidas del formato complejo de los estímulos sonoros procesados con las HRTF medidas.

REFERENCIAS

Algazi, V., Duda, R., Thompson, D., and Avendano, C., The CIPIC HRTF database. IEEE Workshop on applications of Signal Processing to Audio and Acoustics, New Paltz, New York, USA, 99-102, 2001.

Breebaart, J., Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis. J. Acoust. Soc. Am. Express Letters, 133, (1), El1-El6, 2013.

Hu, H., Zhou, L., Ma, H.teh and Wu, Z., HRTF personalization based on artificial network in individual virtual auditory space. Applied Acoustics, 69:163-172, 2007.

Hugeng, Wahidin, W., and Dadang, G., Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis. Signal Processing: An International Journal (SPIJ), 4 (4), 201-212, 2010.

Hölzl, J., An initial Investigation into HRTF Adaptation using PCA. IEM Project Thesis, Institut für elektronische musik und akustik. Graz, Austria, 2012.

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., Isabelle, K,. and Colburn, S., Sensitivity of human subjects to head-related transfer-function phase spectra. J. Acoust. Soc. Am., 105, 5:2821-2840, 1999.

Kulkarni, A. and Colburn, S., Infinite-impulse-response models of the head-related transfer function. J. Acoust. Soc. Am., 115 4:1714-1728, 2004.

Leung and Carlile, C. PCA compression of HRTFs and localization performance, in Proceedings of the International Workshop on the Principles and Applications of Spatial Hearing, 2009.

Mehrgardt, S. and Mellert, V., Transformation characteristics the external human ear, J.
Acoust. Soc. Am. 61:1567–1576, 1977.

Oppenheim A. and Schafer R. Discreet-Time Signal Processing. Prentice-Hall Inc. New Jersey, USA, 1989.

Scarpaci, J., and Colburn S., Principal Components Analysis Interpolation of HRTF's Using Locally Chosen Basis Functions. Proceedings of 11 Meeting of the International Conference on Auditory Display. Limerick, Irlanda, 2005.

Xu, S., Li, Z., and Salvendy G., Identification of Anthropometric Measurements for Individualization of Head-Related Transfer Function. Acta Acustica united with Acustica. 95, 168-177, 2009.