Realidad Virtual Acústica: El Abordaje de las Redes Neuronales Artificiales
Acoustic Virtual Reality: The Artificial Neural Networks Approach

José F. Lucio Naranjo, Roberto A. Tenenbaum y Julio C. B. Torres

o

***Resumen*— En este trabajo se presenta un nuevo abordaje para obtener las respuestas impulsivas biauriculares (BIRs) para un sistema de aurilización utilizando un conjunto de redes neuronales artificiales (RNAs). El método propuesto es capaz de reconstruir las respuestas impulsivas asociadas a la cabeza humana (HRIRs) por medio de modificación espectral y de interpolación espacial. Para poder cubrir todo el espacio auditivo de recepción, sin aumentar la complejidad de la arquitectura de la red, una estructura con varias RNAs (conjunto) fue adoptada, donde cada red opera en una región específica del espacio (gomo). El error de modelaje en el dominio de la frecuencia es investigado considerando la naturaleza logarítmica de la audición humana. A través de la metodología propuesta se obtuvo un ahorro del tiempo de procesamiento computacional de aproximadamente 62% en relación al método tradicional de procesamiento de señales utilizado para aurilización. La aplicabilidad del nuevo método en sistemas de aurilización es reforzada mediante un análisis comparativo de los resultados, que incluyen la generación de las BIRs y el cálculo de un parámetro acústico biauricular (IACF), los cuales muestran errores con magnitudes reducidas.**

***Palabras clave*— Realidad Virtual Acústica. Aurilización. Redes Neuronales Artificiales. Respuestas Impulsivas Bi-auriculares. Simulación Numérica de Acústica de salas.**

***Abstract*— This work presents a new approach to obtain the Binaural Impulse Responses (BIRs) for an auralization system by using a committee of artificial neuronal networks (ANNs). The proposed method is capable to reconstruct the desired modified Head Related Impulse Responses (HRIRs) by means of spectral modification and spatial interpolation. In order to cover the entire auditory reception space, without increasing the network's architecture complexity, a structure with multiple RNAs (committee) was adopted, where each network operates in la specific reception region (bud). The modeling error, in the frequency domain, is investigated considering the logarithmic nature of the human hearing. It was observed that the proposed methodology obtained a computational gain of approximately 62%, in terms of processing time reduction, compared to the classical signal processing method used to obtain auralizations. The applicability of the new method in auralization systems is validated by comparative analysis of the results, which includes the BIR's generation and calculation of one binaural acoustic parameter (IACF), showing very low magnitude errors.**

***Index Terms*— Acoustic Virtual Reality. Auralization. Artificial Neural Networks. Binaural Impulse Responses. Room Acoustics Numeric Simulation.**

This paper was submitted on the LAJC on February 2015
José F. Lucio Naranjo, Departamento de Informática y Ciencias de la Computación, Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador; (emal: jose.lucio@epn.edu.ec)
Roberto A. Tenenbaum, Laboratório de Instrumentación en Dinâmica, Acústica y Vibraciones – LIDAV, Departamento de Modelagem Computacional, Instituto Politécnico, Universidade del Estado del Rio de Janeiro, Nova Friburgo, Brasil; (emal: ratanenbaum@gmail.com)
Julio C. B. Torres, Departamento de Expressão Gráfica, Escola Politécnica, Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brasil. (emal: julio@poli.ufrj.br)

I. INTRODUCCIÓN

EL reciente desarrollo de los sistemas computacionales ha revolucionado la vida del ser humano. Inúmeras aplicaciones informáticas han sido introducidas en las más diversas actividades humanas, que van desde soluciones particulares para problemas específicos, hasta herramientas generales de apoyo en la vida cotidiana de las personas. Entre todas estas, los sistemas de realidad virtual han recibido especial atención los últimos años, por el hecho de ser considerados verdaderas interfaces entre el ser humano y los computadores. La realidad virtual procura recrear la percepción sensorial humana y sus aplicaciones pueden darse en representaciones de recintos, simuladores de vuelo, juegos electrónicos, simuladores en parques de diversiones, entre muchas otras.

En el caso particular de la realidad virtual acústica, el objetivo es la aurilización. Ella consiste en sintetizar un sonido, a partir de datos medidos y/o simulados, capaz de estimular al oyente de tal forma que cause la sensación de inmersión acústica en un ambiente (simulado o real). Tales ambientes pueden ser desde una simple sala de aula, un gran teatro, una planta industrial o hasta un espacio urbano. Esa señal sonora debe ser reproducida a un ser humano en un ambiente libre de reflexiones sonoras como, por ejemplo, en una cámara anecóica (utilizando filtros de atenuación de diafonía), o a través de audífonos ecualizados. De esa forma, puede sintetizarse la sensación de espacialidad del sonido

pudiendo simular la habilidad humana de distinguir la posición de la fuente sonora dentro del ambiente virtual [1].

La llave de la aurilización es la respuesta impulsiva bi-auricular (BIR – del inglés, Binaural Impulse Response) para un determinado par fuente / cabeza humana virtual. Computacionalmente, la BIR es generada a partir de un programa de simulación, él cual, al finalizar la fase de cálculo del modelo del ambiente, tendrá almacenadas información (energética, direccional y tempo de llegada) de todos los frentes de onda que alcanzaron el receptor. El espectro energético de cada una de las frentes de onda incidentes es utilizado para modificar el espectro de un par de respuestas impulsivas asociadas a la cabeza humana (HRIR – del inglés, Head Related Impulse Response), una función para cada oído. El par de HRIRs es escogido de una base de dados, considerando la dirección de incidencia del frente de onda. Sin embargo, debido a la naturaleza discreta de tales bases de datos, es altamente improbable que el frente de onda golpee al receptor en la dirección donde existe una función medida, siendo necesario un procedimiento de interpolación para evitar efectos acústicos no deseados, tales como cliques u otras alteraciones sutiles [2].

Después es necesario adicionar cada una de las HRIRs modificadas en el canal correspondiente (izquierdo y derecho) insertando un atraso correspondiente al tempo de llegada del frente de onda al receptor. Ese proceso (interpolación de HRIRs, modificación con el espectro de la frente de onda y sumatorio con atrasos para montaje de la BIR) es repetido para cada uno de los frentes de onda que alcanzan el receptor. Tales iteraciones constituyen el llamado *método tradicional* (MT) para producir la BIR, el cual, a pesar de los diversos algoritmos de procesamiento de señales disponibles, requieren de costo computacional de procesamiento relativamente alto.
EL presente trabajo presenta una nova abordaje para la generación eficiente de las BIRs aplicadas en procesos de aurilización en sistemas de simulación computacional acústica. La metodología consigue inclusive reducir el costo computacional (i.e. tiempo de procesamiento) para generar HRIRs modificadas, minimizando, al mismo tiempo, cualquier error en el modelaje, de manera que este sea imperceptible al ser humano. En este abordaje, son utilizadas redes neuronales artificiales (RNAs). La ventaja de usar estos sistemas está, principalmente, en la capacidad de las RNAs de adquirir, almacenar y utilizar información experimental con base en el funcionamiento de las neuronas biológicas y en la estructura de procesamiento masivamente paralela del cerebro humano.

Esa estructura consigue que un sistema de RNAs, con una arquitectura correctamente escogida y debidamente entrenada, responda con alta rapidez y precisión, siendo, por tanto, una óptima alternativa para el procesamiento de sistemas de realidad virtual acústica. Una red puede almacenar información relevante de las HRIRs (eliminando la necesidad de bancos de dados e interpolaciones) y, por otro lado, su rapidez de respuesta la credencia como una alternativa eficiente frente al método tradicional de procesamiento de señales, que involucra transformadas de Fourier y otras multiplicaciones complexas.

II. Descripción del Modelo de las RNAS

Según Tebelskis [3], una red neuronal artificial consiste en un número potencialmente alto de unidades de procesamiento simple, llamadas *neuronas artificiales*, las cuales, mediante una red de pesos excitadores o inhibidores (pesos sinápticos), influencian el comportamiento de otras neuronas.

Debido a sus propiedades, las RNAs han sido aplicadas en una amplia variedad de campos. Algunos de los más recurrentes incluyen aplicaciones en interpolación de funciones [4], reconocimiento de patrones [5], predicciones de comportamiento [6] etc. En lo que se refiere a aplicaciones en acústica, las RNAs han sido utilizadas en reconocimiento del habla [3][7] y apenas en los últimos años comenzaron ser utilizadas para la extracción de información relevante de las HRTFs, [8] y para personalización de HRTFs de acuerdo con medidas antropomórficas humanas [9].

Con la debida preparación, las RNAs pueden aprender a generar las salidas/respuestas de un sistema cualquiera (linear o no). Para eso, un conjunto de datos de entrada y sus correspondientes salidas (del sistema que actúa como *profesor*, o sea, que entrena la red) son agrupados en *pares de entrenamiento*. Tales pares son presentados a la red durante un proceso llamado de *aprendizaje supervisado con retropropagación* [10].

En este trabajo, después del proceso de entrenamiento, la metodología propuesta fue capaz de reconstruir las HRIRs modificadas, las cuales eran originalmente generadas por medio de modificación espectral y de interpolación espacial. Con el objetivo de cubrir todo el espacio auditivo de recepción, sin aumentar la complejidad de la arquitectura de la red, una estructura con múltiples RNAs fue adoptada, donde cada red opera en una región específica del espacio (gomo), como muestra la Fig. 1.

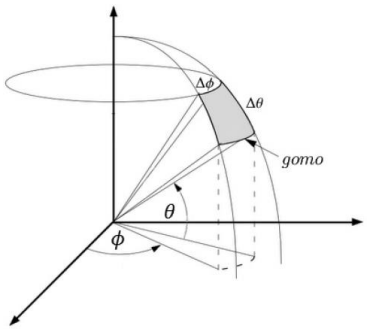

Fig. 1. Área de operación de la red (gomo) limitada en azimut por ϕ y $\phi+\Delta\phi$ y en elevación por θ y $\theta + \Delta\theta$.

Los tres principales factores que influencian en la precisión del modelo — arquitectura de la red, ángulos de abertura del área de recepción y atrasos de las HRIRs — fueron

investigados y una configuración óptima fue determinada [11]. Al final, la arquitectura utilizada en cada una de las redes es mostrada en la Fig. 2. El área de recepción alrededor de la cabeza fue dividida en 1898 gomos y los atrasos de las funciones fueron preservados.

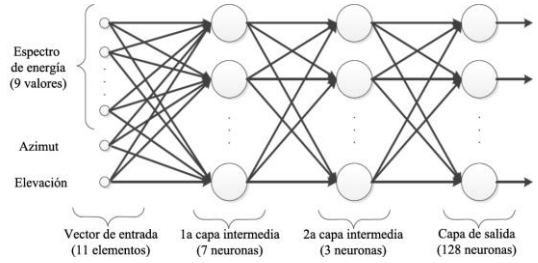

Fig. **2.** Arquitectura básica utilizada por las redes neuronales artificiales con dos capas intermedias.

La entrada de cada RNA está dada por 11 elementos. Dos de ellos representan la dirección de incidencia de las frentes de onda, dada por los ángulos de azimut y elevación. Los nueve valores restantes representan el espectro de energía por banda de octava (63 Hz – 16 kHz). La salida está compuesta por 128 valores, cada uno representando una muestra temporal de las HRIR modificadas. Tales muestras permiten almacenar la información significativa de las funciones [12]. La arquitectura escogida, mostrada en la Fig. 2, tiene dos capas intermedias, la primera con siete neuronas y la segunda con tres neuronas. Esta alternativa presentó el mejor balance entre precisión y costo computacional. Cabe señalar que fue obtenida una ganancia computacional (reducción del tempo de procesamiento) de, aproximadamente, 62% en relación al método tradicional de procesamiento de señales utilizado para aurilización.

III. Resultados Numéricos

El escenario escogido para la validación de los resultados fue la sala utilizada en la comparación internacional de simuladores acústicos Round Robin 3 [13][14]. La mencionada sala fue reconstruida en el código computacional RAIOS (Room Acoustics Integrated and Optimized Software), tal como muestra la Fig 3. El icosaedro rojo representa una fuente sonora omnidireccional activa, el icosaedro gris es una fuente sonora inactiva y las tres esferas verdes representan receptores o cabezas artificiales.

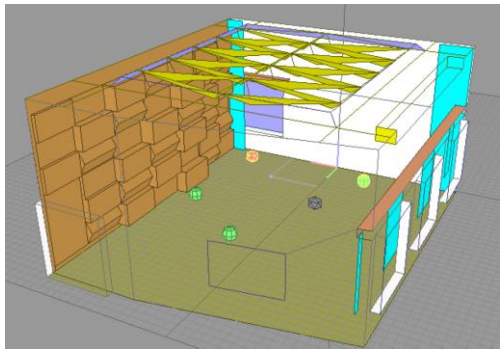

Fig. 3. Estudio de grabación de música en el PTB simulado en el código computacional RAIOS.

En la Fig. 4 pode-se observar la orientación de un de los receptores. EL receptor en cuestión está orientado de manera que el oído derecho está expuesto al sonido directo oriundo de la fuente.

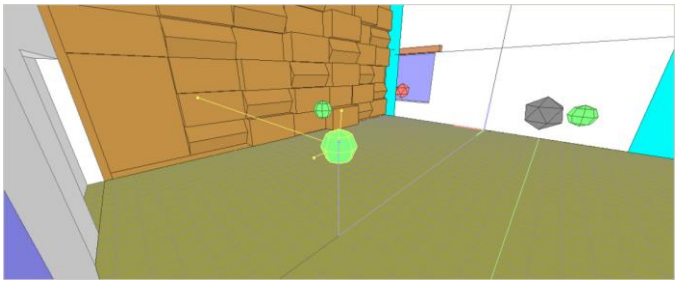

Fig. 4. Orientación de los receptores bi-auriculares.

En primera instancia, la validación fue realizada tomando como referencia los rayos acústicos captados en el receptor durante la simulación acústica, los cuales tuvieron los sus espectros modificados naturalmente pelas características de la propia sala. Con tal información fueron generados resultados con los dos abordajes, el método tradicional (MT) y el método de las redes neuronales artificiales (RNA), relativos al oído derecho.

En la Fig. 5, puede apreciarse el espectro energético de un rayo cuja dirección incidencia es dada pelos ángulos ϕ = 352, 9° y θ = -15, 8°.

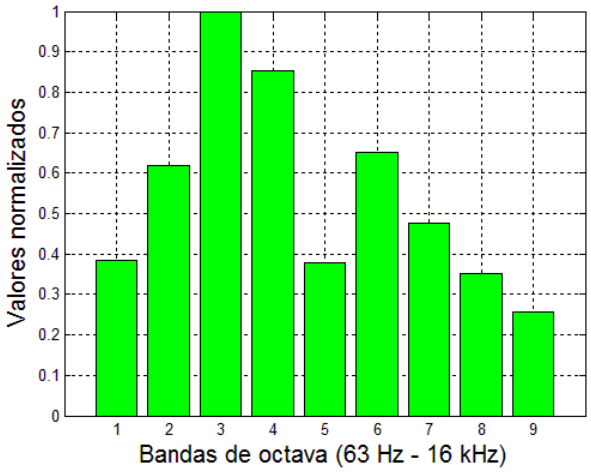

Fig. 5. Espectro de energía del rayo con dirección de llegada (ϕ = 352, 9°, θ = -15,8°).

Puede observarse, en la Fig. 6a, que la función generada por el método RNA, en el dominio del tempo, se adapta a la función generada con el MT con pequeños desvíos indicados en los gráficos de error. En las Figs. 6b y 6d, las diferencias entre las curvas de los espectros de amplitud y fase, obtenidas vía RNAs y el método tradicional, se tornan relevantes solamente a partir de los 7 kHz. Los errores más elevados

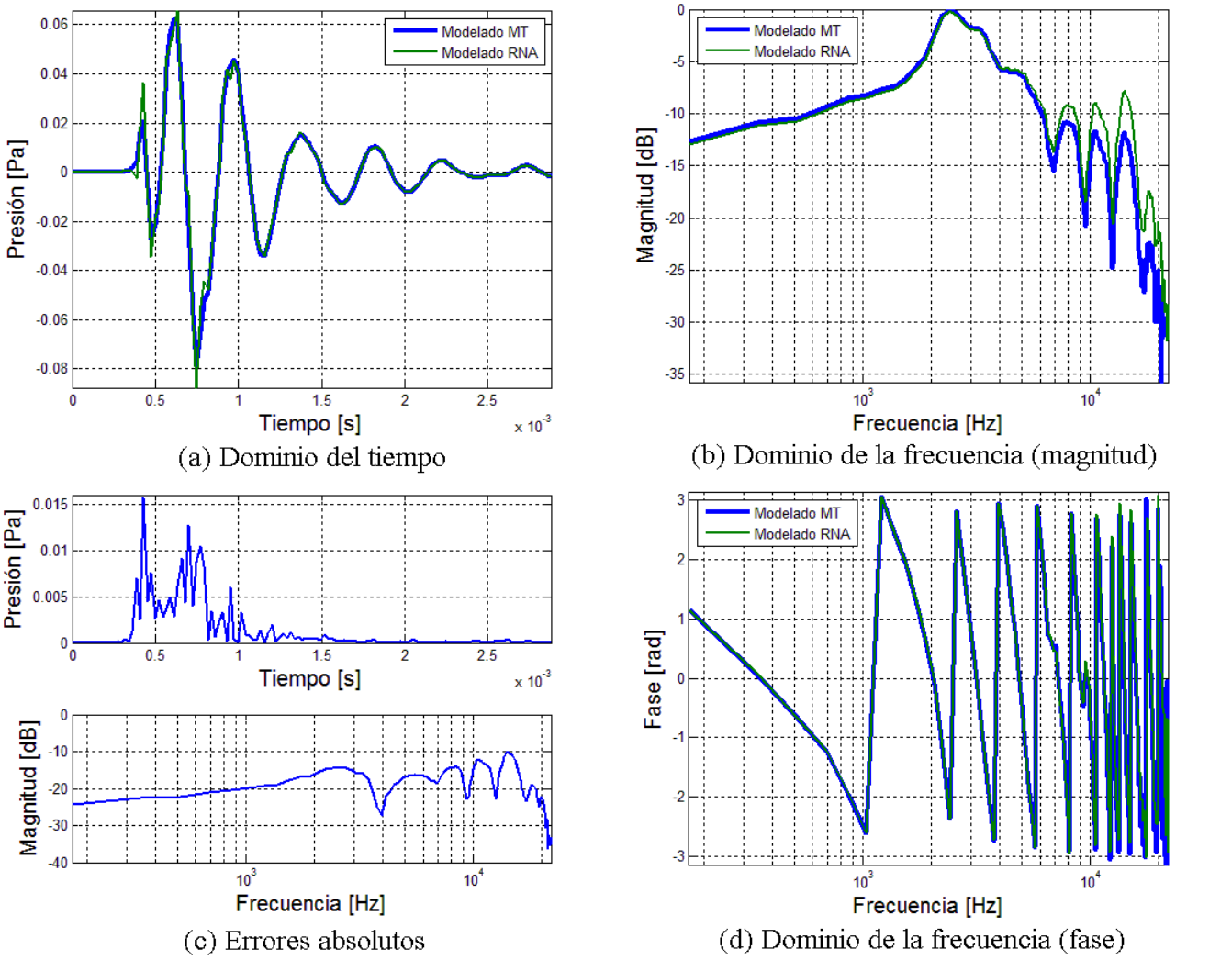

Fig. 6. (a,b,c,d). Resultados generados pelas abordajes RNA y MT, correspondientes al oído derecho, obtenidos para un rayo con dirección dada pelos parámetros θ = -15,8° y ϕ = 352, 9°. En (c), se indican los errores absolutos.

fueron identificados, principalmente en altas frecuencias (encima de 10 kHz), como muestra la Fig. 6. Sin embargo, la mayoría de la energía de las HRIRs modificadas está concentrada en la banda de frecuencia entre 2 y 6 kHz, en la cual el modelo presentó desvíos inferiores la -15 dB. El motivo del desvío está relacionado con la capacidad de generalización de la RNA, la cual, aunque permita generar resultados para entradas que no participaron del proceso de entrenamiento, causa que pequeñas variaciones en las funciones sean desconsideradas, o sea, justamente en las componentes de alta frecuencia.

En una segunda fase de la validación, fue generada la BIR del receptor considerado usando el abordaje de las RNAs. Para ter un resultado comparativo, la BIR también fue generada utilizando el MT. La Fig. 7 presenta los resultados obtenidos. Naturalmente, respuestas impulsivas así presentadas parecen ser gráficamente idénticas.

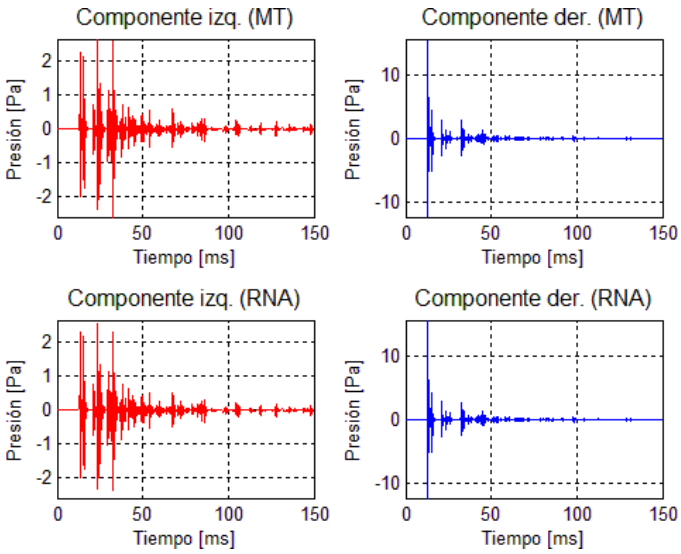

Fig. 7. BIRs generadas con el MT y la técnica de las RNAs.

En la tercera es última validación, fue realizado un test que consideró la correlación existente entre los señales de los dos oídos para poder caracterizar mejor las diferencias entre las BIRs generadas por cada método. Si las funciones de correlación entre los oídos izquierdo y derecho fueran semejantes para ambos métodos, entonces el método RNA habrá preservado las características espaciales y biauriculares, conforme el método tradicional.

En ese sentido fue utilizado el parámetro de calidad acústica bi-auricular conocido como función de correlación cruzada interauricular (IACF - del inglés Inter-Aural Cross-correlation Function)[15]. Los resultados obtenidos son presentados en la Fig. 8.

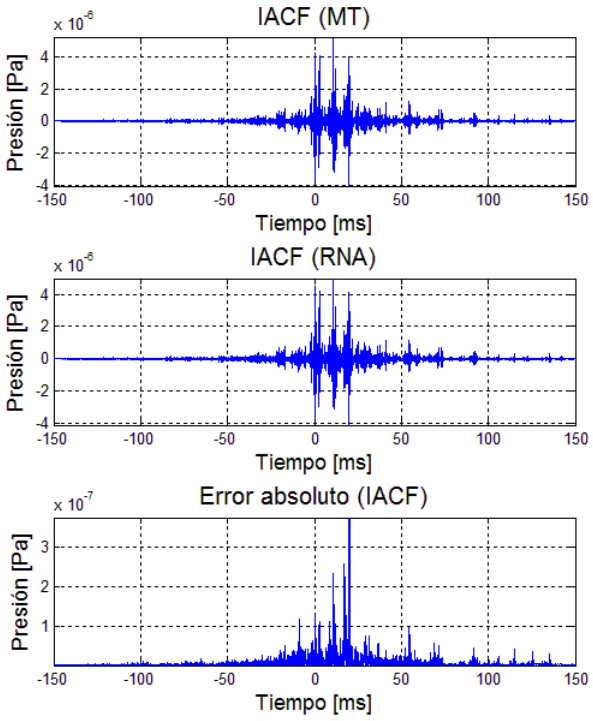

Fig. 8. IACF generadas con las BIRs obtenidas con MT (superior) y la técnica RNA (medio) y el error absoluto entre ellas (inferior).

EL gráfico superior muestra la función generada con la BIR obtenida con el MT, el gráfico intermedio presenta los resultados obtenidos con la técnica RNA y el gráfico inferior presenta el error absoluto entre el IACFs del MT y del RNA. Tales resultados están en concordancia con la orientación que favorece el oído derecho, tanto con el MT como con las RNAs. Los máximos valores alcanzados en lo que se refiere al error absoluto para el IACF tienen una orden de grandeza de 10^{-7}.

IV. Conclusiones

Fue presentado un nuevo abordaje para la generación eficiente de las respuestas impulsivas biauriculares (BIRs). En ese intuito, un conjunto de redes neuronales artificiales (RNAs) fue utilizado para substituir procesos de interpolación y la modificación espectral aplicados a las respuestas impulsivas asociados a la cabeza humana (HRIRs).

Los resultados numéricos mostraron que el procesamiento de señales tradicional puede ser reemplazado por un conjunto de RNAs cuya salida presenta errores próximos a -17 dB en relación al método tradicional. Tanto los errores absolutos entre las funciones, como los resultados comparativos utilizando el parámetro de calidad acústica biauricular IACF, muestran que un conjunto de redes, cada una con una arquitectura reducida (apenas dos capas intermedias con 7 y 3 neuronas respectivamente), es capaz de reemplazar al método tradicional, computacionalmente desventajoso.

Los resultados obtenidos a partir de rayos acústicos generados pelo simulador acústico RAIOS, los cuales no fueron considerados durante entrenamiento, muestran que la generalización de la red funciona con errores de la orden de -15 dB. Sin embargo, el precio de una buena capacidad de generalización es la concentración de los desvíos en altas frecuencias.

En la práctica, las BIRs generadas con el método tradicional o con las RNAs no presentan diferencias significativas. El parámetro biauricular IACF presentó resultados que estaban en concordancia con los parámetros escogidos para la simulación. Más aún, el error absoluto calculado entre los IACFs generados pela técnica clásica y la metodología propuesta alcanzan valores de la orden de 10^{-7}. Eso confirma numéricamente la precisión del modelo propuesto.

References

[1] BLAUERT, J., 1997. Spatial Hearing. The MIT Press, Cambridge.
[2] WERSNYI, G., 2009. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation IEEE Transactions On Audio, Speech, And Language Processing, Vol. 17, n. 2, pp. 247–252.
[3] TEBELSKIS, J., 1995. Speech Recognition using Neuronal Networks. Ph.D. Thesis - Carnegie Mellon University, Pittsburgh.
[4] MINGO, L. F.; GIMÉNEZ, V.; CASTELLANOS, J., 1999 Interpolation of boolean functions with enhanced neuronal networks. In: Second Conference on Computer Science and Information Technologies. CSIT'99. Yerevan, Armenia. p. 17–22.
[5] BISHOP, C., 2005. Neuronal Networks for Pattern Recognition. Oxford: Oxford University Press.
[6] ALMEIDA, F., PASSARI, A., 2006. Aplicação de redes neurais na previsão de vendas no varejo. Revista de Administração - RAUSP, v. 41, n. 3, p. 257–272.
[7] HOLMES, J., HOLMES, W., 2001. Speech Synthesis and Recognition. 11 New Fetter Lane, London: Taylor & Francis. 213–218 p.
[8] HARASZY, Z., IANCHIS, D., TIPONUT, V., 2009. Generation of the head related transfer functions using artificial neuronal networks. 13th WSEAS International Conference on CIRCUITS, p. 114–118.
[9] HU, H.; ZHOU, L.; MA H. & WU, Z., 2008. HRTF personalization based on artificial neuronal network in individual virtual auditory space. Applied Acoustics, v. 69, n. 2, p. 163–172.
[10] RUMELHART, D., MCCLELLAND, J., the PDP Research Group, 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, 328–330 p.
[11] LUCIO NARANJO, J.F.L., 2014. Inteligência Computacional Aplicada na Geração de Respostas Impulsivas Biauriculares e em Aurilização de Salas. Ph.D. Thesis, Universidade do Estado do Rio de Janeiro, Nova Friburgo.
[12] TORRES, J.C.B., PETRAGLIA, M.R., AND TENENBAUM, R.A., 2004. An Efficient wavelet-based HRTF for auralization Acta Acustica united with Acustica, Vol. 90(1), 108–120.

[13] TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND GERGES, S.N.Y., 2007a. Hybrid method for numerical simulation of room acoustics with auralization: Part 1 - Theoretical and numerical aspects. J. Brazilian Soc. Mech. Sci. Engin., Vol. 29(2), 211–221.
[14] TENENBAUM, R.A., CAMILO, T.S., TORRES J.C.B. AND STUTZ, L.T., 2007b Hybrid method for numerical simulation of room acoustics: Part 2 - Validation of the computational code RAIOS 3. J. Brazilian Soc. Mech. Sci. Engin., Vol. 29(2), 222–231.
[15] VORLÄNDER, M., 2008. Auralization: Fundamentals of Acoustics, Modeling, Simulation, Algorithms and Acoustic Virtual Reality Springer, Berlin

José Francisco Lucio Naranjo nació en Quito - Ecuador, el 01 de junio de 1979. Obtuvo el grado de ingeniero en Sistemas Informáticos por la Pontificia Universidad Católica del Ecuador (2005). En 2010 y en 2014, obtuvo, respectivamente, sus grados de M.Sc. y de Ph.D, en Modelado Computacional en el IPRJ, Universidad del Estado de Río de Janeiro, Brasil. Su principal campo de estudio está enfocado en técnicas de simulación numérica acústica aplicadas en la generación de realidad virtual aplicando redes neuronales artificiales.

El Dr. Lucio está actuando como investigador y profesor titular de informática en la Escuela Politécnica Nacional del Ecuador (EPN). También ha actuado como profesor sustituto en la Universidad del Estado de Río de Janeiro (UERJ), tutor presencial en el Programa de Posgrado en Educación a Distancia en la Universidad Federal Fluminense, profesor curricular en la Universidad de las Américas (UDLA) y profesor auxiliar en la Universidad Central del Ecuador (UCE). También ha sido miembro asociado de la Sociedad Brasileña de Acústica (SOBRAC).

Roberto A. Tenenbaum es Ingeniero Mecánico por la Universidad Federal de Río de Janeiro, en 1972; obtuvo su M.Sc. en el campo de Termoelasticidad por la COPPE, Universidad Federal de Río de Janeiro, en 1975, y su D.Sc. en Ciencias en el campo de la acústica en la misma universidad, en 1987. Su principal campo de interés es la dinámica, acústica y vibraciones.

Ha trabajado como ingeniero en la Compañía de Tecnología Nuclear, en Río de Janeiro, de 1973 a 1974. Luego, como profesor e investigador en el Laboratorio de Acústica y Vibraciones (LAVI) perteneciente al Programa de Postgrado de Ingeniería Mecánica, en la Universidad Federal de Río de Janeiro, RJ-Brasil, de 1974 a 2004. Actualmente, se desempeña como profesor e investigador en el Laboratorio de Instrumentación en Dinámica, Acústica y Vibraciones (LIDAV) en el Programa de Postgrado en Modelado Computacional, en la Universidad de Estado de Río de Janeiro, en Nova Friburgo, Brasil. Es autor de tres libros sobre dinámica: Dinâmica, 1997, en portugués; Fundamentals of Applied Dynamics, 2004, en Inglés, publicado por Springer-NY; y Dinâmica Aplicada, 2006, también en portugués. Él es también el autor y coautor de más de 150 artículos publicados en revistas y congresos.

El Profesor Tenenbaum es miembro honorario y fundador de la Sociedad Brasileña de Ciencias Mecánicas e Ingeniería, ABCM; un miembro fundador de la Sociedad Brasileña Acústica, SOBRAC; Es miembro de la Acoustical Society of America, ASA; y la Sociedad Internacional de problemas inversos en Ciencias e Ingeniería, ISIPSE, entre otros.

Julio Cesar Boscher Torres nació en Río de Janeiro, Brasil, el 20 de diciembre de 1971. Obtuvo su grado de Ingeniero Eléctrico en 1993 y, en 1998 y 2004, respectivamente, obtuvo su M.Sc. y su Ph.D., en Ingeniería Eléctrica, de la Universidad Federal de Río de Janeiro (COPPE / UFRJ). Tiene experiencia en Ingeniería Eléctrica, que abarca los siguientes temas: procesamiento de señales, ondas, aurilización, acústica de salas y simulación acústica. Otras áreas de interés son la música, grabación en estudio, stereophotogrametry y visión 3D y simulación. Desde 2004, trabaja en la UFRJ como profesor asociado, impartiendo cátedra e investigando en las áreas mencionadas. El Prof. Torres es un miembro de IEEE.