

Descripción y Análisis de un Modelo de Localización Sonora Espacial

R. Venegas C¹., R. Correa² y S. Floody³

^{1,3}Departamento de Acústica, Universidad Tecnológica de Chile, Brown Norte 290, Ñuñoa, Santiago, Chile.

²Departamento de Física, Universidad Tecnológica Metropolitana, José Pedro Alessandri 1242, Ñuñoa, Santiago, Chile.

¹rodolfo.venegas@gmail.com, ²rcorrea@utem.cl, ³sergio.floody@utc.cl

Resumen

En este trabajo se presenta la descripción y análisis de un modelo computacional de localización sonora espacial para sonidos de banda ancha en ambiente anecoico inspirado en el sistema auditivo humano e implementado mediante técnicas de inteligencia computacional. El modelo es un transformador no lineal que posee un módulo de extracción de características significativas: diferencias interaurales de tiempo, nivel y espectrales de nivel y un segundo módulo constituido por una red neuronal artificial feedforward de tres capas que entrega como resultado la estimación de la posición de la fuente sonora en los ángulos de elevación y azimut. Un análisis de sensibilidad de las entradas del modelo y su importancia relativa también son presentados.

1. INTRODUCCIÓN

La capacidad de reconocer desde donde proviene un sonido y eventualmente asociar la fuente sonora que lo emite a una posición en el espacio, muchas veces resulta ser de vital importancia, a dicha habilidad de los seres vivos se le denomina localización sonora espacial. En las últimas décadas la localización sonora espacial en seres vivos ha sido estudiada en gran profundidad [1-6] obteniendo importantes avances en la comprensión del sistema auditivo y en como éste decodifica las pistas que permiten la localización. A su vez, con el aumento de la capacidad computacional, la implementación de modelos cuyo fin es comprender de mejor forma o modelar el proceso de localización sonora espacial se mantiene como un fecundo campo de investigación [7-17].

La localización sonora espacial es un problema que involucra un conocimiento en muchos ámbitos y cuya modelación aun no ha podido ser implementada con total éxito. Desde el punto de vista biológico, uno de los principales motivos es que cada persona posee una fisiología del oído externo distinta y por ende una función de transferencia relacionada a la cabeza (HRTF) diferente [18-21], lo cual dificulta la caracterización del problema en sí. Por otra parte, otro trascendental motivo es que aún no ha sido comprendida totalmente en seres vivos.

El uso de técnicas de Inteligencia Computacional, tales como redes neuronales artificiales (RNA) y su capacidad de aprendizaje, las sitúan como una poderosa herramienta para resolver el problema en cuestión [7-15]. En este trabajo se presenta un modelo computacional de localización sonora espacial para sonidos de banda ancha en ambiente anecoico inspirado en el sistema auditivo humano e implementado mediante el uso de técnicas de inteligencia computacional. El proceso de construcción del modelo y los resultados de dicho proceso han sido presentados en [15] [22-26], selectos resultados de tales referencias y un análisis de sensibilidad y relevancia de las entradas del modelo también son presentados. El trabajo está organizado de la siguiente forma. En la sección 2 se indica el procedimiento de medición; en la sección 3 se describe el modelo y sus componentes; en la sección 4 se discute los resultados obtenidos por el modelo y se realiza un análisis de sensibilidad y relevancia de sus entradas. Finalmente, en la sección 5 se concluye.

2. PROCEDIMIENTO DE MEDICIÓN.

El sistema de referencia usual para el estudio de localización sonora espacial es en coordenadas esféricas y se presenta en la figura 1. En la figura 2 se presentan los planos cuyos vectores normales corresponden a los ejes cartesianos presentados en la figura 1, estos planos son usualmente referidos en la literatura como plano medio, frontal y horizontal [1-2] [27]. En este trabajo se hace uso del denominado sistema coordenado vertical polar [27] y se considera que el

ángulo de elevación φ varía desde -90° a 90° (abajo hacia arriba) y el de azimut θ de 0° a 360° (derecha a izquierda).

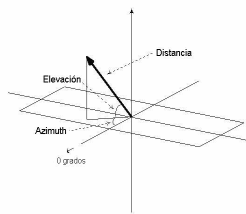


Fig. 1. Sistema de Referencia

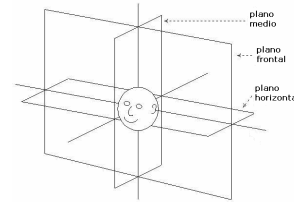


Fig. 2. Plano medio, frontal y horizontal.

Para el entrenamiento del modelo de localización sonora espacial es necesario poseer mediciones experimentales. Es común encontrar en la literatura trabajos en los cuales se realiza sólo la medición de las HRTF y luego se obtiene el conjunto de entrenamiento y el de prueba mediante la convolución de señales con las HRIR (Head Related Impulse Responses) [4] [7] [10] [12] [16]. En este trabajo no se sigue dicho procedimiento sino que se realiza la adquisición de datos para la construcción del conjunto de entrenamiento y prueba de forma netamente experimental ya que dicho proceso permite obtener una representación más adecuada de las condiciones auditivas normales. El proceso de medición se realizó en base a [28-30] y consistió en situar, a 1,2 [m] del origen del sistema de referencia, un parlante en una estructura especialmente construida y medir con una cabeza artificial de prueba (CAP) -colocada en el centro de la cámara anecoica de la Universidad Tecnológica de Chile- sonidos sweep logarítmicos de banda ancha sin énfasis de 1 s y ruidos blancos, generados con diferentes programas, de 100 ms, 250 ms y 500 ms en pasos de 5° en azimut para 0° elevación, pasos de 10° en azimut para elevaciones entre -60° y 60° (separadas en pasos de 15°) y pasos de 20° en azimut para 75° de elevación. Adicionalmente, se midieron otras direcciones aleatoriamente. Un total de 1708 mediciones fueron utilizadas para confeccionar los conjuntos de entrenamiento y prueba. Para más detalles véase [22-23] [26].

3. DESCRIPCIÓN DEL MODELO

El modelo propuesto es un transformador no lineal cuyas entradas son sonidos de banda ancha estacionarios captados por el oído derecho e izquierdo de la CAP y su salida es la estimación de la posición espacial de la fuente sonora en los ángulos de azimut y elevación. Dicho modelo posee dos módulos. El primero realiza -mediante procesamiento de las señales- la extracción de las siguientes características: diferencias interaurales de tiempo (ITD), de nivel (ILD) y espectrales de nivel (ISLD). El segundo módulo está constituido por una red neuronal artificial (RNA) que estima la posición espacial de la fuente sonora. La figura 3 presenta un esquema del modelo propuesto. En las sub secciones de este apartado se describen los componentes del modelo.

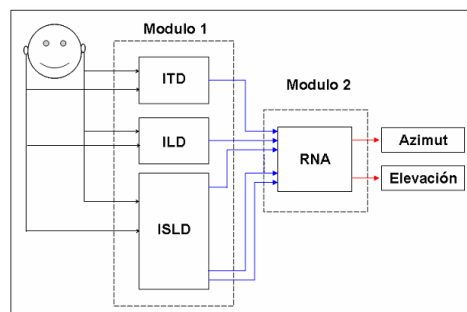


Fig. 3. Esquema del Modelo de Localización Sonora Espacial

3.1 Diferencia Interaural de Tiempo ITD

La diferencia interaural de tiempo ITD se define como la diferencia entre el tiempo de arribo de una señal a cada oído. En este trabajo el método para calcular ITD se basa en el propuesto por Jeffress

[3], éste hace uso de la correlación cruzada entre las señales de ambos oídos ya que el vector de correlación cruzada es proporcional a ITD [3]. Las ecuaciones (1) y (2) describen el cálculo de ITD, en ellas $x_R[n]$ es la señal del oído derecho, $x_L[n]$ la señal del oído izquierdo, N es la longitud de ambas señales en muestras y $\text{nint}(\cdot)$ es la función entero más cercano. La figura 4 presenta ITD en función del ángulo de azimut (abscisa) y de elevación (ordenada).

$$C_{x_R x_L}[m] = \begin{cases} \sum_{n=0}^{N-m-1} x_R[m+n]x_L[n] & m \geq 0 \\ C_{x_R x_L}[-m] & m < 0 \end{cases} \quad (1) \quad ITD = \arg \max_m (C_{x_R x_L}[m]) - \text{nint}\left(\frac{2N-1}{2}\right) \quad (2)$$

3.2 Diferencia Interaural de Nivel ILD

La diferencia interaural de nivel ILD se define como la diferencia entre los niveles de las señales que arriban a cada oído. En este trabajo ILD se calcula como la razón entre las energías de las señales como indica (3). La figura 5 presenta ILD en función de azimut y elevación.

$$ILD = 10 \cdot \log\left(\sum_n (x_R[n])^2\right) - 10 \cdot \log\left(\sum_n (x_L[n])^2\right) \quad (3)$$

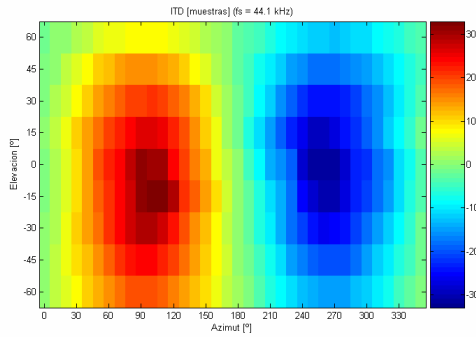


Fig. 4. ITD calculado a partir del sweep logarítmico

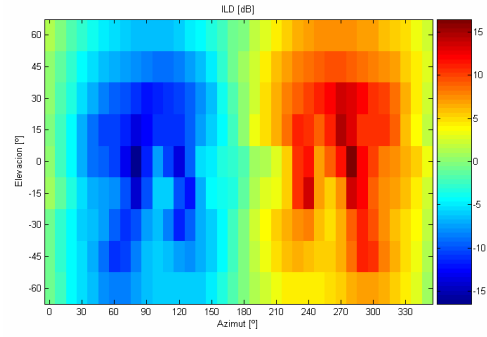


Fig. 5. ILD calculado a partir del sweep logarítmico

3.3 Diferencias Interaurales Espectrales de Nivel ISLD

Las diferencias interaurales espectrales de nivel ISLD se definen como las diferencias de energía por banda de frecuencia entre las señales -filtradas con un banco de filtros- que arriban a ambos oídos. El algoritmo para el cálculo de ISLD es descrito por las ecuaciones (4) a (8).

$$X^{(R,L)}(f) = 20 \log |FFT(x_{(R,L)}[n])| \quad (4) \quad X_s^{(R,L)}(f) = \frac{1}{BW_s} \int_{f-\frac{BW}{2}}^{f+\frac{BW}{2}} X^{(R,L)}(f_p) df_p \quad (5)$$

$$Y_{(k,F)}^{(R,L)}(f) = X_s^{(R,L)}(f) + 20 \log |BF_{(k,F)}(f)| \quad (6) \quad E_F^{(R,L)}[k] = 10 \log \left(\frac{1}{BW_{(k,F)}} \int_{f_L(k,F)}^{f_H(k,F)} 10^{0.1 Y_{(k,F)}^{(R,L)}(f)} df \right) \quad (7)$$

$$ISLD_F[k] = E_F^{(R)}[k] - E_F^{(L)}[k] \quad (8)$$

A partir de (4) se obtiene la magnitud del espectro de la señal mediante el cálculo de una

Transformada rápida de Fourier FFT que permita una resolución espectral de 1 Hz. Con (5) se suaviza la magnitud del espectro mediante una media móvil; si la suavización no es abrupta, esta no influye en el proceso de localización sonora espacial [31], para este trabajo $BW_s = 160$ Hz, cabe señalar que $f = f_0 : f_1$; $f_0 > 80$ Hz y $f_1 < 21,97$ kHz. Utilizando (6) se filtran en magnitud los espectros suavizados de cada oído con un banco de filtros BF que posee k filtros. En [15] [25-26] se hace uso de tres bancos de filtros (ERB, Lyon y de Tercio de Octava). Los mejores resultados se obtienen utilizando 18 filtros ERB extraídos del modelo auditivo propuesto por Patterson y Holdsworth (Equivalent Rectangular Bandwidth Patterson Holdsworth Auditory Filter Bank) [32-34], en la figura 6 se presenta la magnitud de la respuesta en frecuencia de dicho banco de filtros. Se emplearon filtros con frecuencias centrales mayores a 1,25 kHz ya que el propósito de este proceso es ofrecer información significativa para la estimación del ángulo de elevación [1-2]. Empleando (7) se calcula la energía por banda $E_F^{(R,L)}$ para cada oído; cada banda posee frecuencias de corte inferior $f_L(k, F)$ y superior $f_H(k, F)$. Mediante (8) se calcula ISLD como la diferencia entre las energías por banda de la señal derecha y la izquierda.

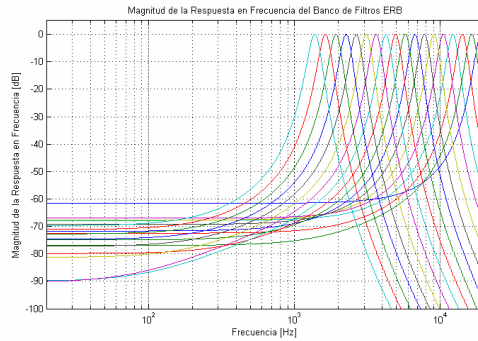


Fig. 6. Magnitud de la Respuesta en Frecuencia Banco de Filtros ERB

3.4 Vector de Características

Se aplicaron (1) a (8) a cada elemento de la base de datos. Como salida del primer módulo se obtienen los vectores de características z , dados por (9), dependientes de la posición espacial de la fuente. Para ISLD calculado con el banco de filtros ERB, $D = 20$. Las componentes de ISLD se ordenaron en función de mayor a menor frecuencia central.

$$z(\theta, \varphi)_{BF} = [ISLD_{BF} \quad ILD \quad ITD]_{D \times 1} \quad (9)$$

3.5 Redes Neuronales Artificiales RNA

Un completo desarrollo teórico y práctico de redes neuronales artificiales se puede encontrar en [35] y [36]. Las entradas a la RNA corresponden a los vectores de características normalizados con el promedio de los máximos valores absolutos que tomaron estos considerando todos los tipos de sonido. Las salidas son los ángulos de azimut y elevación normalizados al intervalo $[0,1]$ y $[-1,1]$ respectivamente. La configuración de la RNA que brindó mejores resultados [15] [25-26], corresponde a una red neuronal artificial feedforward con 20 neuronas en la capa de entrada, 15 en la capa oculta y 2 en la de salida (20-15-2), con funciones de activación tangente hiperbólica sigmoide (tansig), logarítmica sigmoide (logsig) y función lineal (purelin), la cual fue entrenada con el algoritmo Quasi-Newton Backpropagation considerando 5000 épocas, un error objetivo igual a $1e-5$, gradiente mínimo $1e-6$ y paso mínimo de $1e-6$. La determinación de esta configuración responde a un proceso de diseño, presentado en [15] [25-26].

4. RESULTADOS Y DISCUSIÓN

El modelo óptimo fue hallado mediante un procedimiento introducido en [15] y depurado en [25-26]. Básicamente, el procedimiento consiste en hallar las configuraciones que minimizan el error en el conjunto de entrenamiento, prueba y entrenamiento y prueba como un solo conjunto y calcular un indicador de sobre entrenamiento y generalización [26] en orden de obtener valores numéricos que permitan determinar objetivamente la configuración óptima. Como ya se había mencionado, la configuración óptima corresponde a una red feedforward de 3 capas 20-15-2 con funciones de transferencia tansig-logsig-purelin, cuyas entradas fueron ITD, ILD e ISLD calculadas utilizando el banco de filtros ERB. Dicha configuración fue entrenada con 1433 elementos y probada con 275 (cada elemento está asociado a una dirección espacial). El error -medido con la distancia euclídea promedio- en el conjunto de entrenamiento es 1.92° , en el de prueba 4.3° y de 2.3° en el de entrenamiento y prueba como un solo conjunto.

En cuanto a la cantidad de entradas: el modelo de Anderson et al. [7] posee como entradas las diferencias espectrales de fase o de potencia filtradas con bancos de filtros de tercio de octava o doceava de octava (al menos 48 entradas). Backman y Karjalainen [8] proponen un modelo con 69 entradas. Grassi y Shamma [10] emplean 69 filtros cocleares para el cálculo de ITD e ISLD (denominado ILD en ese trabajo, 138 entradas en total). Jin et al. [12] utilizan 62 entradas. Neti et al. [13] extraen 128 muestras de las HRTF normalizadas como entradas. Las entradas del modelo de Steinberg et al. [14] son los cosenos de las fases de N detectores teóricos; encuentran que el número mínimo de detectores es 31, por lo cual el número de entradas es el mismo. Martin [16] ocupa 24 filtros de tercio de octava a partir de los cuales extrae diferencias interaurales de nivel, tiempo y fase (72 entradas). Macpherson [37] ocupa 120 filtros e ITD (121 entradas). Recientemente, Park et al. [38] presentan un modelo que utiliza 60 filtros gammatone para el cálculo de ITD e ILD por banda (120 entradas) que sirven de entrada a un sistema de decisión. Si bien algunos de los modelos citados no sólo son para sonidos de banda ancha, el modelo propuesto posee una menor cantidad de entradas que todos ellos y estas revelan, en base a los resultados y el análisis de sensibilidad, ser necesarias y suficientes.

Si se considera la topología del modelo: Backman y Karjalainen [8] utilizan RNA 2-4, 4-4 y 4-8 obteniendo los mejores resultados con estas últimas. Datum et al. [9] ocupan una RNA 32-32-32. Jin et al. [12] emplean una RNA MLP (perceptron multicapa) 40-20-393 y unas RNA TDNN (Time delay neural network) 40-20-393, 60-20-393 y 50-28-393 obteniendo los mejores resultados para sonidos de banda ancha con la RNA MLP. Neti et al. [13] hacen uso de topologías con 128 neuronas en la capa de entrada, 4 a 10 en la capa oculta y un arreglo de 11×17 (azimut x elevación) neuronas en la capa de salida. Nuevamente tomando en consideración que algunos modelos no sólo son para sonidos de banda ancha, el modelo propuesto en este trabajo presenta una topología con un menor número de parámetros, salvo el presentado en [8].

Referente al error del modelo: Backman y Karjalainen [8] utilizaron ruidos blancos y rosas entre otros para construir el conjunto de entrenamiento; la RNA fue probada con el conjunto de entrenamiento como "conjunto de prueba" obteniendo un error promedio de 0.2° en el plano horizontal, cuando entrenaron la red para detectar azimut y elevación el error promedio fue de 1° . Grassi y Shamma [10] utilizando sólo 69 direcciones, obtiene un error RMS de 0 metros para ruidos blancos. Jin et al. [12] para sonidos de banda ancha en mediciones a 76 sujetos obtienen un error medio de ángulo esférico de 11° ; utilizando redes MLP 40-20-393, TDDN 40-20-393, TDDN 60-20-393, TDDN 50-28-393 y TDDN 50-28-393 con conjunto de entrenamiento construido aleatoriamente obtiene un error medio de ángulo esférico de 5° , 9° , 10° , 9° y 7° respectivamente. Neti et al. [13] obtienen con su modelo binaural y conjunto de entrenamiento construido aleatoriamente un error de 0.3° en el conjunto de entrenamiento y de 6.3° en el de prueba. El modelo de Martin [16] posee un error de 8° promedio para sonidos de banda ancha. Searle et al. [17] resumen los resultados de experimentos realizados en humanos a partir de los cuales confeccionaron su modelo, de dichos experimentos se aprecia un error entre 0° y 0.24° para el plano horizontal y entre 3.6° y 20.2° para el plano medio utilizando sonidos de banda ancha. Buttler [39] realizó mediciones a 20 sujetos, los cuales obtuvieron un error de 0.24° en el plano

horizontal y $3,6^\circ$ en el plano vertical para sonidos de banda ancha. Hebrank y Wright [40] en mediciones a 10 sujetos obtiene un error de $12,3^\circ$ en el plano vertical para sonidos de banda ancha. Se aprecia una gran disparidad en los resultados y el modelo propuesto en este trabajo posee un error dentro de los márgenes encontrados en la literatura.

Para cada tipo de sonido (sweep y ruidos blancos) se realizó la propagación hacia adelante de los vectores de entrada que representan direcciones en azimut de 0° a 350° en pasos de 10° y de -60° a 60° en pasos de 15° para elevación. A partir de las salidas, se calculó la distancia euclídea entre la respuesta de la red y los ángulos objetivos y se promediaron para cada dirección con respecto al tipo de sonido. La figura 7 muestra dicho resultado. Se aprecia que el modelo distingue con una precisión óptima si la fuente sonora se encuentra adelante o atrás (front-back confusion) o si se encuentra arriba o abajo. Por otra parte, los errores no presentan un patrón espacial aparente, situación que el modelo no comparte con la localización sonora espacial en humanos [41].

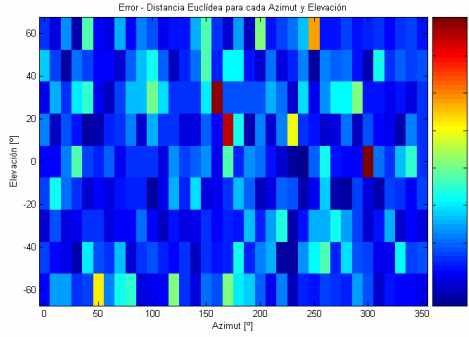


Fig. 7. Distancia euclídea para Azimut y Elevación - Error por Dirección

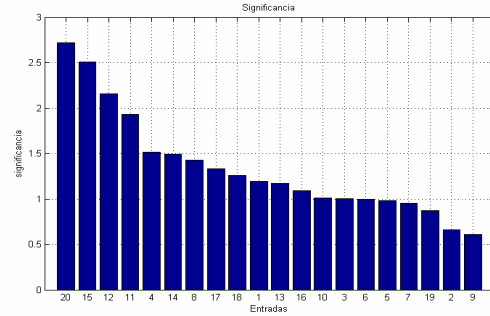


Fig. 8. Significancia de cada Entrada.

En orden de cuantificar la influencia de cada entrada del modelo se realizó un análisis de sensibilidad de la mejor configuración, el cual se describe a continuación. Sea $W_{(I \times D)}$, $V_{(H \times I)}$ y $U_{(O \times H)}$; $B_{(I \times I)}^{[In]}$, $B_{(H \times I)}^{[Hi]}$ y $B_{(O \times I)}^{[Ou]}$ y f_{In} , f_{Hi} y f_{Ou} las matrices de peso, bias y funciones de activación de la capa de entrada (*In*), oculta (*Hi*) y salida (*Ou*) respectivamente. La salida de la RNA del modelo de localización sonora espacial propuesto en este trabajo está dada por (10) y las matrices de sensibilidad [42] para cada patrón de entrada p se definen mediante (11).

$$r = f_{Ou} \left(U f_{Hi} \left(V f_{In} \left(Wz + B^{[In]} \right) + B^{[Hi]} \right) + B^{[Ou]} \right) \quad (10) \quad S_{kd}^{(p)} = \frac{\partial r_k}{\partial z_d} \quad (11)$$

Aplicando la regla de la cadena se puede obtener la expresión matricial explícita de las matrices de sensibilidad dada por (12). En donde $C^{[In]}$, $C^{[Hi]}$ y $C^{[Ou]}$ son matrices diagonales cuyas entradas se calculan como la variación de la respectiva función de activación con respecto a el valor de entrada a dicha función, por ejemplo $C^{[In]}$ se calcula mediante (13).

$$S^{(p)} = C^{[Ou]} \times U \times C^{[Hi]} \times V \times C^{[In]} \times W \quad (12) \quad C_{(I \times I)}^{[In]} = \text{diag} \left(\frac{df_{In}(Wz + B^{[In]})}{d(Wz + B^{[In]})} \right) \quad (13)$$

$$S_{avg} = \sqrt{\frac{1}{N_T} \sum_{p=1}^{N_T} (S^{(p)})^2} \quad (14) \quad \hat{S}_{avg}^{kd} = S_{avg}^{kd} \cdot \frac{\left(\max_{n=1 \dots N_T} \{z_d^{(n)}\} - \min_{n=1 \dots N_T} \{z_d^{(n)}\} \right)}{\left(\max_{n=1 \dots N_T} \{r_k^{(n)}\} - \min_{n=1 \dots N_T} \{r_k^{(n)}\} \right)} \quad (15)$$

Para determinar la influencia de cada entrada se deben calcular las matrices de sensibilidad sobre todo el conjunto de entrenamiento y el de prueba como un solo conjunto. En orden de obtener un valor de sensibilidad para cada componente de las entradas en [42] se propone una métrica denominada matriz de sensibilidad cuadrática media promedio S_{avg} dada por (14). Donde N_T es el número total de elementos del conjunto de entrenamiento y el de prueba como un solo conjunto. Si las entradas y salidas no fueron escaladas a los mismos valores es necesario un post-procesamiento, dado por (15), el cual permite comparar adecuadamente la sensibilidad de cada entrada. La significancia de cada entrada [42] se calcula utilizando (16). La figura 8 muestra la significancia de cada entrada (abscisa) ordenada de mayor a menor.

$$\Phi_{d,avg} = \max_{k=1...O} \{ \hat{S}_{kd,avg} \} \quad (16)$$

La entrada más significativa corresponde a ITD{20} situación que también se presenta en seres vivos [1-2] [27]. La segunda entrada más relevante es ISLD{15} (frecuencia central $f_c=2265$ Hz) la cual corresponde a la banda donde se produce la frecuencia de resonancia del “canal auditivo” de la CAP, la tercera y cuarta corresponden a ISLD{12} ($f_c=3635$ Hz) e ISLD{11} ($f_c=4242$ Hz), en tales frecuencias centrales la pina manifiesta su mayor influencia y brinda valiosa información para la determinación de elevación [1-2] [27]. En base a los resultados se puede afirmar que el modelo muestra características que se presentan en la localización sonora espacial humana. Las entradas con menor significancia corresponden a ISLD{9} ($f_c=5758$ Hz), ISLD{2} ($f_c=16407$) e ILD{19}. Es posible que la baja influencia de ILD se deba a algún fenómeno similar a la multicolinealidad en las regresiones lineales ya que ILD e ITD están correlacionados fuertemente, no obstante, dicha interpretación y el resultado en sí, merecen un estudio más detallado. Adicionalmente, en [42] se indica un criterio para eliminar las entradas con menor significancia, debido a que el error no disminuye al eliminar las entradas con menor significancia (2 y 9) [25], se concluye que las 20 entradas del modelo óptimo son necesarias para obtener el desempeño exhibido.

5. CONCLUSIONES

En este trabajo se presenta una descripción y el análisis de un modelo computacional de localización sonora espacial para sonidos de banda ancha en ambiente anecoico inspirado en el sistema auditivo humano e implementado mediante redes neuronales artificiales. En función de los resultados se concluye que el modelo es más sencillo y eficiente que algunos propuestos en la literatura [7], [9] [12], [13], [14], [16], [37], [38]; que exhibe características que se presentan en la localización sonora espacial humana y que, en base al análisis de sensibilidad, las 20 entradas del modelo revelan ser necesarias y suficientes en orden de obtener un error de 4,3 %dirección en el conjunto de prueba. Las principales limitaciones del modelo son su particularización a sonidos de banda ancha y que sólo ha sido testeado en ambientes anecoicos, trabajos futuros serán realizados en dichos tópicos. Potenciales aplicaciones son: evaluación y calibración de imágenes sonoras virtuales en sistemas de reproducción sonora multicanal, adhesión de un modelo a sonoridad direccional y fusión con sistemas de visión computacional.

REFERENCIAS

- [1] J. Blauert, “Spatial Hearing, The Psychophysics of Human Sound Localization”, Cambridge MA: MIT Press, 1983.
- [2] D. Begault, “3-D Sound for Virtual Reality and Multimedia”, NASA Ames Research Center, Moffet Field California, 2000.
- [3] L. Jeffress, “A place theory of sound localization” J. Comp. Physiol. Psychol. Vol. 41, pp. 35–39, 1948.
- [4] D. Nandy and J. Ben-Arie, “Neural models for auditory localization based on spectral cues”, Neurological Research, Vol. 23, No. 5, pp. 489-500, Jul. 2001.
- [5] M. Van Wanrooij and J. Van Opstal, “Contribution of Head Shadow and pinna Cues to Chronic Monoaural Sound Localization”, Journal of Neuroscience, Vol. 24, No 17, pp. 4163-4171, April 2004.
- [6] M. Van Wanrooij and J. Van Opstal, “Relearning Sound Localization with new Ear”, Journal of Neuroscience, Vol. 25, No. 22, pp. 5413-5424, June 2005.
- [7] T.R. Anderson, J. A. Janko and R.H. Gilkey, “An Artificial neural network model of human sound”, J. Acoust. Soc. Am., Vol. 92, No. 4, pp. 2298, Oct. 1992.

- [8] J. Backman and M. Karjalainen, "Modelling of Human directional and Spatial Hearing using neural networks", IEEE ICASSP 1993, Vol. 1, pp.125-128, Minneapolis, Minnesota, USA, 1993.
- [9] M. S. Datum, F. Palmieri and A. Moiseff, "An artificial neural network for sound localization using binaural cues" J. Acoust. Soc. Am. Vol. 100, No.1, pp. 372-383, 1996.
- [10] E. Grassi and S. Shamma, "A biological inspired, learning, sound localization algorithm", Conference on Information Sciences and Systems, The Johns Hopkins University, March 21-23, 2001.
- [11] J. A. Janko, T. Anderson and R. Gilkey "Using neural networks to evaluate the viability of monaural and interaural cues for sound localization" in Binaural and Spatial Hearing in Real and Virtual Environments, edited by R. Gilkey, T. Anderson and L. Erlbaum, Mahwah, Chap. 26, pp. 557-570, 1997.
- [12] C. Jin, M. Schenkel and S. Carlile, "Neural system identification model of human sound localization" J. Acoust. Soc. Am. Vol. 108 No. 3 Pt. 1., p. 1215-1235, 2000.
- [13] C. Neti, E. Young and M. Schneider, "Neural network model of sound localization based on directional filtering by the pinna", J. Acoust. Soc. Am., Vol. 92, No. 6, pp. 3140-3156, Dec.1992.
- [14] B. Steinberg, M. Beran, S. Chin, and J. Howard, "A neural network approach to source localization", J. Acoust. Soc. Am. Vol. 90, No. 4 Pt. 1, pp. 2080-2090, 1991.
- [15] R. Venegas, M. Lara, R. Correa and S. Floody, "Spatial Sound Localization Model Using Neural Network", 120th AES Convention, Paris, France, May 19-23, 2006.
- [16] K. Martin, "A computational model of Spatial Hearing", MSc. dissertation, Massachusetts Institute of Technology MIT, 1995.
- [17] C. Searle, L. Braida, D. Cuddy and M. Davis, "Model for auditory localization" J. Acoust. Soc. Am., Vol. 60, No. 5, pp.1164-1175, 1976.
- [18] V. Algazi, P. Diventi and R. Duda, "Subject dependent Transfer Function in Spatial Hearing", IEEE MWSCAS 1997, Vol. 2, pp.877-880, 1997.
- [19] S. Carlile and D. Pralong, "The location-dependent nature of perceptually salient features of the human related transfer function", J. Acoust. Soc. Am., Vol. 95 No. 6, pp. 3445-3459, 1994.
- [20] E. M. Wenzel, F.L. Wighman, D.J. Kistler and S. H. Foster, "Acoustics origin of individual differences in sound localization behavior", J. Acoust. Soc. Am., Vol 84, No. 1, p. S79, Nov. 1988.
- [21] E. M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wighman, "Localization using nonindividualized head related transfer function", J. Acoust. Soc. Am., Vol. 94, No 1, pp. 111-123, July 1993.
- [22] R. Venegas, M. Lara, R. Correa y S. Floody, "Diseño, Implementación y Entrenamiento de un Sistema de Identificación Sonora Espacial Neuro-Difuso", Reporte Técnico Proyecto FONDEI, Universidad Tecnológica de Chile, 2005.
- [23] R. Venegas, M. Lara, R. Correa y S. Floody, "Medición de HRTF y extracción de características significativas para localización sonora", Seminario Internacional de Acústica SEMACUS 2005, Universidad Pérez Rosales, Santiago, Chile, 27-29 Octubre 2005.
- [24] R. Venegas, M. Lara, R. Correa, S. Floody, "Modelo de Localización Sonora Espacial 2D para Sonidos de Banda Ancha", EMFIMIN 2005, Universidad de Santiago, 10 - 11 de Noviembre, 2005..
- [25] R. Venegas, R. Correa y S. Floody, "Modelo Computacional de Localización Sonora Espacial", Ingelectra 2006, Universidad Austral de Chile, Valdivia, Chile, 23-25 Agosto 2006.
- [26] R. Venegas, "Diseño e implementación de un modelo de localización sonora espacial utilizando técnicas de inteligencia computacional", Tesis conducente al título de Ingeniero Civil en Sonido y Acústica, Universidad Tecnológica de Chile, Santiago, Chile, 2006.
- [27] R. O. Duda, "3-D Audio for HCI", Department of Electrical Engineering, San Jose State University, Available: http://interface.cipic.ucdavis.edu/CIL_tutorial/3D_home.htm.
- [28] V. Algazi, R. O. Duda, D. Thompson and C. Avendano, "The CIPIC HRTF Database", Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.
- [29] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy Head Microphone", MIT Media Lab Perceptual Computing, Technical Report No 280, May 1994.
- [30] H. Moller, M. Sorensen, D. Hammershoi and C. Jensen, "Head-Related Transfer Function of Human Subjects", J. Audio Eng Soc., Vol. 43, No 5, pp. 300-321, 1995.
- [31] A. Kulkarni and H.S. Colburn, "Role of spectral detail in sound-source localization", Nature, Vol. 396 (6713), pp. 747-749; Dec. 24-31, 1998.
- [32] R.D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and H. Allerhand, "Complex sounds and auditory images" in Auditory Physiology and Perception, (Eds) Cazals, Y., Demany, K., and Horner, K., Pergamon, Exford, 1992.
- [33] M. Slaney, "An efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank", Apple Computer Technical Report No. 35, Advanced Technology Group, Cupertino. CA., 1993.
- [34] M. Slaney, "Auditory Toolbox v. 2", Technical Report #1998-010, Interval Research Corporation, 1998.
- [35] M. Gupta, L. Jin and N. Homma, "Static and Dynamic Neural Network", IEEE Press, John Wiley and Sons, 2003.
- [36] M. T. Hagan, H. B. Demuth and M. H. Beale, "Neural Network Design", Boston, MA: PWS Publishing, 1996.
- [37] E. A. Macpherson, "A comparison of spectral correlation and local feature-matching models of pinna cue processing" J. Acoust. Soc. Am. Vol. 101, No. 5 part 2, pp. 3104 (A), 1997.
- [38] M. Park, P. Nelson and Y. Kim, "An auditory process model for the evaluation of virtual acoustic imaging system", 120th AES Convention, Paris, France, May 2006.
- [39] R. Butler, "Monoaural and binaural Localization of Noise Burst Vertically in the Median Sagittal Plane", J. Aud. Res. Vol. 3, pp. 245-246., 1969.
- [40] J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?". J. Acoust. Soc. Am., Vol 56, No 3, pp. 935-938, 1974.
- [41] S. Carlile, P. Leong and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," Hearing Research 114, pp. 179-196. 1997.
- [42] M. Zurada, A. Malinowski, and I. Cloete, "Sensitivity analysis for minimization of input data dimension for feedforward neural network". In Proceedings of the IEEE Symposium on Circuits and Systems, pp. 447-450, 1994.