

PRIMERAS JORNADAS REGIONALES DE ACÚSTICA AdAA 2009

19 y 20 de noviembre de 2009, Rosario, Argentina



AdAA2009-A028R

Análisis comparativo entre modelos de sonoridad: propuesta de un modelo de ponderación dinámica

Alejandro Osses Vecchi^(a)
Víctor Espinoza Catalán^{(a), (b)}

(a) Departamento de Sonido y Acústica, Universidad Tecnológica de Chile, INACAP. Brown Norte 290, Ñuñoa, Santiago, Chile. E-mail: osses.vecchi@yahoo.com

(b) Centro Tecnológico, Facultad de Artes, Universidad de Chile. Compañía 1264, Santiago, Chile. E-mail: vespinozac@gmail.com

Abstract

The reference standard ISO 532:1975 is commonly used for calculating the approximated loudness level for complex acoustic signals. The so-called Method B proposes an analysis based on 1/3 octave bands for the approximation to the critical bands of the human hearing system. In this paper an alternative model is proposed which uses the warping technique to obtain a critical band frequency domain and a dynamic weighting model implemented by means of an Artificial Neural Network (ANN). The model is compared with the ISO 532-B method and with the Patterson-Holdsworth auditory model, which is based on the concept of Equivalent Rectangular Bandwidth (ERB).

Resumen

La norma de referencia ISO 532:1975 es comúnmente utilizada para el cálculo aproximado del nivel de sonoridad para señales acústicas complejas. El denominado Método B plantea el análisis en bandas de un tercio de octava para la aproximación a las bandas críticas del sistema auditivo humano. En el presente trabajo se propone un modelo de sonoridad alternativo, usando la técnica de *warping* de frecuencias para obtener un dominio en bandas críticas y un modelo de ponderación dinámico implementado mediante una Red Neuronal Artificial (RNA). La propuesta es comparada con el método ISO 532-B y con el modelo auditivo de Patterson y Holdsworth, este último basado en el concepto de *Equivalent Rectangular Bandwidth* (ERB).

1 Introducción

Las características de sensibilidad y de selectividad del oído humano como función de la frecuencia son variables altamente no lineales. La consideración de estas características para el análisis de señales acústicas se adecua mejor a la valoración perceptual en comparación a la utilización de metodologías que consideran variables netamente físicas como el filtrado en bandas de tercios de octava y el nivel de presión sonora en dB. Pese a lo anterior, existen 2 métodos normalizados que utilizan el análisis en bandas de octava y/o en tercios de octava para la estimación de la sonoridad de señales acústicas estacionarias. En el presente trabajo se propone una metodología alternativa para el cálculo de la sonoridad en condición de campo libre, basado en un banco de 24 filtros en bandas críticas además de las curvas de igual sonoridad (ISO 226:2003).

2 Bandas Críticas

El concepto de bandas críticas define el funcionamiento del sistema auditivo humano como un conjunto de filtros pasabanda, de manera tal que 2 bandas críticas con igual intensidad sonora presentarán la misma sonoridad.

El concepto de banda crítica está asociado al fenómeno psicoacústico de enmascaramiento. Una banda crítica se define como un filtro pasabanda que contiene todas aquellas componentes útiles de un ruido de banda ancha que contribuyen en el enmascaramiento de un tono de prueba centrado en una frecuencia determinada.

El sistema auditivo humano es modelado usualmente como un banco de 24 bandas críticas (Zwicker, 1961), cada una de las cuales estimula una igual cantidad de células ciliadas de la membrana basilar.

La transducción de ondas mecánicas a impulsos eléctricos es realizada en la membrana basilar (MB) donde a lo largo de sus 32 mm se distribuyen las mencionadas células ciliadas. Dependiendo el punto de estímulo en la MB, es la altura o frecuencia del impulso que se percibirá. Por este motivo, es posible relacionar la frecuencia f en Hz con el punto de estímulo x en mm en la MB, sin embargo esta relación es no lineal.

En (Zwicker y Terhardt, 1980) se introdujo analíticamente la escala Bark de frecuencias, que presenta una relación aproximadamente lineal respecto al punto de estímulo en la MB y está definida por la denominada razón de banda crítica z , de acuerdo a la ecuación (1).

$$z = 13 \times \arctan(0,76 \times f) + 3,5 \times \arctan\left(\frac{f}{7,5}\right)^2, \quad (1)$$

donde f es la frecuencia en kHz y la razón de banda crítica z es medida en Barks.

Los valores enteros adquiridos por la razón de banda crítica z corresponden a las frecuencias de corte –propuestas en (Zwicker, 1961)– de cada una de las bandas críticas que modelan el oído humano, por lo que el ancho de banda crítico (CB_c) en Barks es constante, de 1 Bark. El ancho de banda crítico CB_c expresado en Hz se puede determinar como función de la frecuencia central de cada banda crítica f_c utilizando la ecuación (2).

$$CB_c = 25 + 75 \times \left[1 + 1,4 \times f_c^2\right]^{0,69}, \quad (2)$$

donde f_c está expresada en kHz.

La ecuación (2) puede aproximarse considerando que CB_c corresponde al 20 % de f_c para frecuencias sobre 500 Hz¹.

2.1 Equivalent Rectangular Bandwidth, ERB

Una escala de frecuencias alternativa es la escala ERB (*Equivalent Rectangular Bandwidth*), la cual surge al aproximar un filtro auditivo a través de una banda perfectamente rectangular, en función de su frecuencia central f_c . El ancho de banda ERB del filtro auditivo en Hz está dado por la ecuación (3).

$$ERB = 0,108 \times f_c + 24,7 . \quad (3)$$

El ancho de banda ERB corresponde aproximadamente a un 11% de su frecuencia central, por lo que los filtros auditivos equivalentes poseen un ancho de banda menor que el presentado por las bandas críticas (20% de f_c sobre 500 Hz). Debido a ello, se requiere una mayor cantidad de filtros ERB para caracterizar el rango completo de frecuencias del sistema auditivo humano (de 20 Hz a 22.050 Hz).

La escala auditiva de frecuencias ERB está definida por la ecuación (4).

$$ERB_N = 21.4 \times \log_{10}(0,0047 \times f + 1) , \quad (4)$$

donde f es la frecuencia en Hz y ERB_N es el número ERB (razón ERB).

2.2 Filtros auditivos de Patterson-Holdsworth

En (Patterson et al, 1992) es descrito un banco de filtros que modela el funcionamiento coclear del oído humano, basándose en una implementación optimizada de los llamados filtros Gammatone en el dominio ERB de frecuencias. El modelo se encuentra implementado en el *Auditory Toolbox 2* para MATLAB® y está disponible en la web². La figura (1) muestra algunas bandas pertenecientes al banco de filtros de Patterson-Holdsworth.

2.3 Aproximación a la escala de banda crítica: Técnica de *Warping* de frecuencias

Esta técnica consiste en la aplicación de una transformación bilineal a un filtro digital. Esta transformación depende de un parámetro λ conocido como factor de *warping*. El mapeo de frecuencias del dominio digital lineal (ω) al dominio digital con *warping* (ν) está dado por:

$$\nu(\omega) = \omega + 2 \times \arctan\left(\frac{\lambda \times \sin(\omega)}{1 - \lambda \times \cos(\omega)}\right) . \quad (5)$$

Como nuestro objetivo es obtener una relación lineal entre el nuevo dominio digital ν y la razón de banda crítica z , debe encontrarse un factor de *warping* apropiado. Para ello se utiliza la ecuación (6) (Smith y Abel, 1999)

¹ Para frecuencias inferiores a 500 Hz, el ancho de banda crítico CB_c es considerado constante, de 100 Hz.

² Descarga de *Auditory Toolbox 2* para MATLAB®: <http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/>

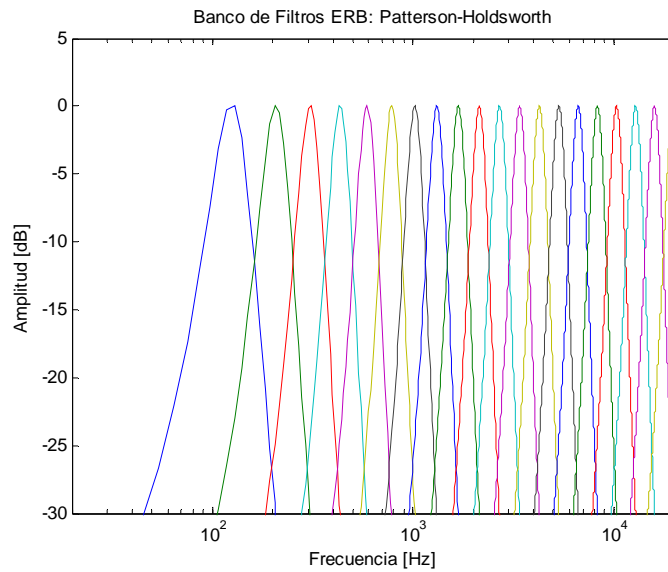


Figura 1. Algunos filtros ERB de acuerdo al modelo coclear de Patterson-Holdsworth.

$$\lambda(f_s) = 1,0674 \times \left[\frac{2}{\pi} \times \arctan(0,06583 \times f_s) \right]^{\frac{1}{2}} - 0,1916, \quad (6)$$

donde f_s es la frecuencia de muestreo de la señal de audio en Hz. El factor de *warping* λ considerando un muestreo de 44.100 muestras por segundo corresponde a 0,7564.

La frecuencia digital ν varía entre 0 y π , por lo que la aproximación a la razón de banda crítica z estará dada por:

$$z_\nu = \frac{\nu}{\pi} \times z_s, \quad (7)$$

donde z_s es la frecuencia de muestreo expresada en Barks usando la ecuación (1). De esta forma, la escala z_ν será considerada equivalente a la razón de banda crítica z .

2.4 Aproximación de las Bandas Críticas: Modelo de enmascaramiento de Terhardt

Para la aproximación de las pendientes inferiores y superiores de cada banda crítica, se hará utilización del Modelo de Enmascaramiento de Terhardt (Terhardt et al, 1982):

$$\begin{aligned} S_1 &= 27 & f < F_M \\ S_2 &= -24 - 230 / F_M + 0,2 \times SPL & f > F_M \end{aligned} \quad (8)$$

donde S_1 y S_2 están expresados en dB/Bark, F_M representa la frecuencia del tono enmascarador en Hz, que coinciden con las frecuencias centrales de las bandas críticas definidas en (Zwicker, 1961) y L_p es el nivel de presión sonora en dB.



Figura 2. Modelo de Terhardt usado para la aproximación de una banda crítica.

Cada banda crítica tendrá un ancho de banda aproximado de 1 Bark y su implementación digital se efectuará disponiendo un filtro pasabajos y un filtro pasa altos en cascada (Osse y Espinoza, 2009).

3 Modelos de Sonoridad

3.1 Sonoridad: definición

La sonoridad o *loudness* es una medida subjetiva de la intensidad de un sonido, de manera que 2 tonos puros con igual sonoridad son percibidos a igual intensidad por el oído humano. Se define el nivel de sonoridad L_s , como el nivel de presión sonora (L_p) a 1 kHz que es percibido a igual intensidad que un L_p a alguna determinada frecuencia. La unidad empleada para el nivel de sonoridad es el fon (*Phon*). El análogo subjetivo a la presión sonora es la sonoridad, medida en Sones (*Sone*) y está definida por la ecuación (9) (Bladon, 1981):

$$S = \begin{cases} 2^{\frac{(L_s - 40)}{10}} & \text{Si } L_s \geq 40 \\ \left(\frac{L_s}{40}\right)^{2,642} & \text{Si } L_s \leq 40 \end{cases} \quad (9)$$

De la ecuación (9), se puede obtener la expresión analítica para el nivel de sonoridad L_p , en Fones:

$$L_s = \begin{cases} 10 \times \log_2(S) + 40 & \text{Si } S \geq 1 \\ 40 \times S^{0,3785} & \text{Si } S < 1 \end{cases} \quad (10)$$

La relación entre L_p en dB y el nivel de sonoridad L_s en Fones se encuentra rotulada en la norma ISO 226:2003, donde se presentan las curvas de igual sonoridad o curvas isofónicas, sin embargo, esta relación está condicionada a tonos puros y no a señales acústicas complejas.

Por otro lado, existen 2 métodos para el cálculo de la sonoridad de una señal acústica compleja, que se encuentran estandarizados en la norma ISO 532:1975. De allí, se hará utilización del método conocido como Método B, que considera un análisis en bandas críticas estimado a partir de un análisis en bandas de tercios de octava.

3.2 Sonoridad Total S y Sonoridad Específica S'

Si se tienen dos tonos de prueba de igual nivel, en la medida en que la separación de las frecuencias de ambos tonos aumenta, la sensación de sonoridad total también aumenta. Debido a lo anterior, resulta útil pensar la sonoridad total S como la integral de la denominada Sonoridad Específica S' (Son/Bark), dependiente de la razón z (Barks), (Zwicker y Fastl, 2007). La ecuación (11) presenta la relación entre S y S' y su aproximación discreta.

$$S = \int_0^{24 \text{ Bark}} S' dz \approx \sum_j S'_j \Delta z_j, \tag{11}$$

donde S es la sonoridad total, en sones, Δz_j corresponde a la distancia en Barks entre la frecuencia z_{j-1} y z_j y S'_j es la sonoridad específica asociada a z_j

3.3 Contornos de igual sonoridad

Los contornos de igual sonoridad o curvas isofónicas representan la respuesta del oído humano ante diferentes niveles y frecuencias. Las curvas fueron determinadas empíricamente utilizando tonos puros y considerando una incidencia normal en el observador.

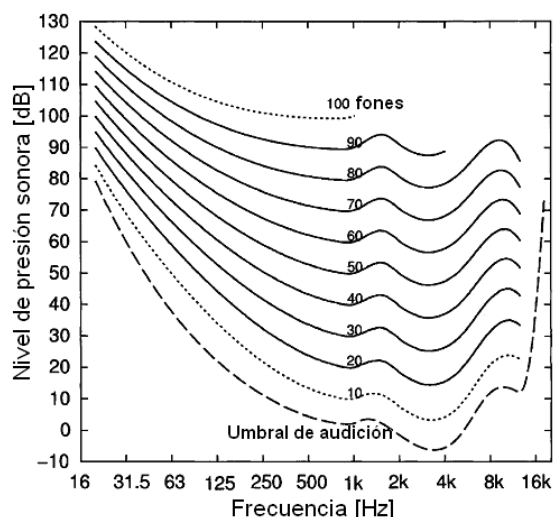


Figura 3. Contornos de igual sonoridad (Suzuki y Takeshima, 2004).

Las curvas de la figura (3) corresponden a los contornos de igual sonoridad estandarizados en la norma ISO 226:2003, las cuales muestran discrepancias de más de 10 dB en algunas regiones, principalmente de baja frecuencia, con respecto a la derogada ISO 226:1987.

3.3.1 Redes Neuronales Artificiales: Aproximación funcional de las curvas isofónicas

En (Espinoza et al, 2006) se implementó una Red Neuronal Artificial del tipo *Feed-Forward* para la aproximación funcional de las curvas isofónicas. En el presente trabajo se hace utilización de esta red para la transformación de nivel de presión sonora L_p a nivel de sonoridad L_s .

En general, una red neuronal artificial RNA es una estructura que emula el funcionamiento de un conjunto de neuronas humanas teniendo, por tanto, la capacidad de aprender a relacionar sus entradas con sus salidas para, más tarde, responder de manera autónoma frente a entradas similares a las usadas en el denominado proceso de entrenamiento.

Se usó una red *Feed-Forward* de 3 capas (2 capas ocultas y 1 capa de salida), con 11-9-1 neuronas y funciones de transferencia tangente sigmoideal, logarítmica sigmoideal y lineal, respectivamente. Las entradas a la red fueron la frecuencia en Hz y el nivel de presión sonora L_p en dB, mientras que la salida fue el nivel de sonoridad L_p en Fones. Durante el proceso de entrenamiento se determinan los valores de los pesos sinápticos de la red, así como los bias asociados a cada neurona.

3.4 ISO 532-B

También conocido como Método B, considera los contornos de igual sonoridad así como las características de transmisión del oído externo y medio hacia el oído interno. Como aproximación a las bandas críticas del oído humano utiliza las bandas de un tercio de octava. La principal desventaja de este método es que está estandarizado como un procedimiento gráfico en el cual a partir del trazado de los niveles L_p de cada banda de tercio de octava se obtiene la sonoridad específica (Sonoridad por unidad de frecuencia: Son/Bark), para luego calcular la sonoridad total como el área bajo la curva trazada.

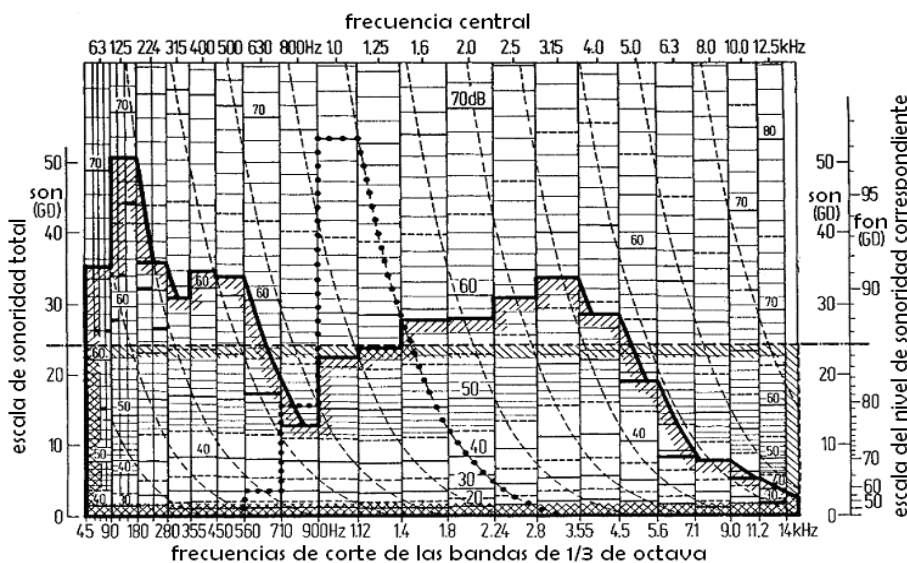


Figura 4. Cálculo de sonoridad usando el Método B (ISO 532:1975-B). Imagen extraída de (Zwicker y Fastl, 2007).

En (Zwicker et al, 1991) se describe un algoritmo computacional que permite concretar el cálculo de sonoridad específica y sonoridad total de acuerdo a este procedimiento gráfico³.

4 Modelo propuesto

Se propone un modelo para el cálculo de sonoridad en condición de campo libre, basado en un análisis en el dominio de la frecuencia. Posee las siguientes etapas:

³ Más tarde, A. Hastings y P. Davies (Universidad de Purdue) desarrollarían un script de MATLAB® conteniendo este algoritmo.

- a. Aplicación de una ventana temporal a la señal de audio de entrada.
- b. Conversión del dominio temporal al dominio de la frecuencia.
- c. Filtrado de la señal usando un banco de filtros en bandas críticas aproximado por la técnica de *Warping* de frecuencias.
- d. Cálculo del enmascaramiento global de la señal de audio por cada banda crítica. Se usa el modelo de Terhardt considerando una separación de 17 dB entre el tono enmascarador con su respectiva banda enmascarada.
- e. Ponderación dinámica de la señal usando la RNA ya entrenada, almacenada en la función de MATLAB® NeuLoud. Conceptualmente, NeuLoud entrega el nivel de sonoridad específico L'_s (Fon/Bark).
- f. El nivel de sonoridad específico L'_s es convertido a sonoridad específica (Son/Bark), usando la ecuación (9) en cada uno de los valores de L'_s .
- g. Cálculo de la sonoridad por cada banda crítica: área bajo la curva de sonoridad específica: utilización de la ecuación (11).
- h. Cálculo de la sonoridad total: suma aritmética de la sonoridad de cada una de las 24 bandas críticas.

Cada elemento del diagrama en bloques de la figura (5) representa cada una de las etapas del método recién descrito.

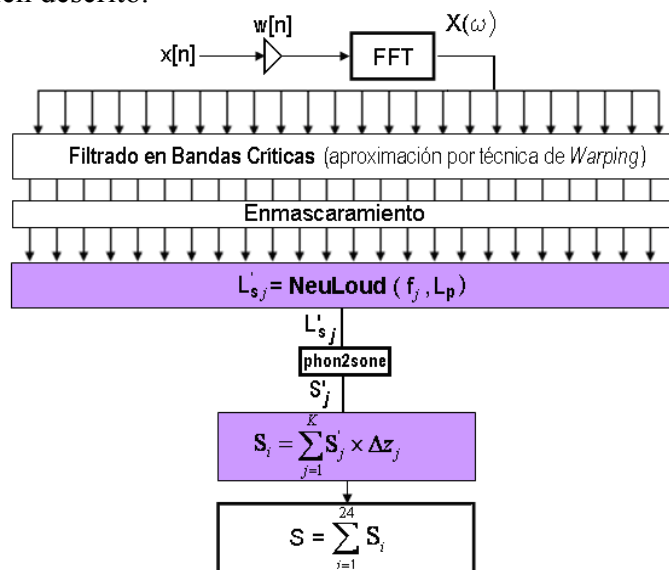


Figura 5. Modelo propuesto para el cálculo de la sonoridad de una señal digital. Se esquematizan las etapas descritas en los puntos (a-f).

5 Ejemplos

Se determinó la sonoridad en condición de campo libre de dos fragmentos de audio, referidos como `Audio1` y `Audio2`, almacenados en formato wav con una frecuencia de muestreo de 44.100 Hz, 16 bits. Además, se usó una ventana del tipo Chebyshev con una longitud de 2.048 muestras.

Para el proceso de calibración se consideró como referencia un L_p de 70 dB, ajustando el nivel cuadrático medio de las señales digitales a este valor.

El cálculo de sonoridad de acuerdo al Método B, se implementó considerando el algoritmo propuesto en (Zwicker et al, 1991), considerando campo libre. Para efectos de comparación, se realizó una suma de la sonoridad de cada banda crítica⁴. Posteriormente, estos valores fueron convertidos a nivel de sonoridad, en Fones, usando la ecuación (10).

El modelo propuesto, se realiza implementando el sistema de la figura (5). El banco de filtros en bandas críticas ha sido diseñado utilizando la técnica de *warping* de frecuencias como aproximación a la escala Bark de frecuencias (ver sección 2.3).

El modelo de Patterson-Holdsworth ha sido implementado reemplazando las 24 bandas críticas de la figura (5) por los 42 filtros auditivos ERB que caen dentro del rango auditivo.

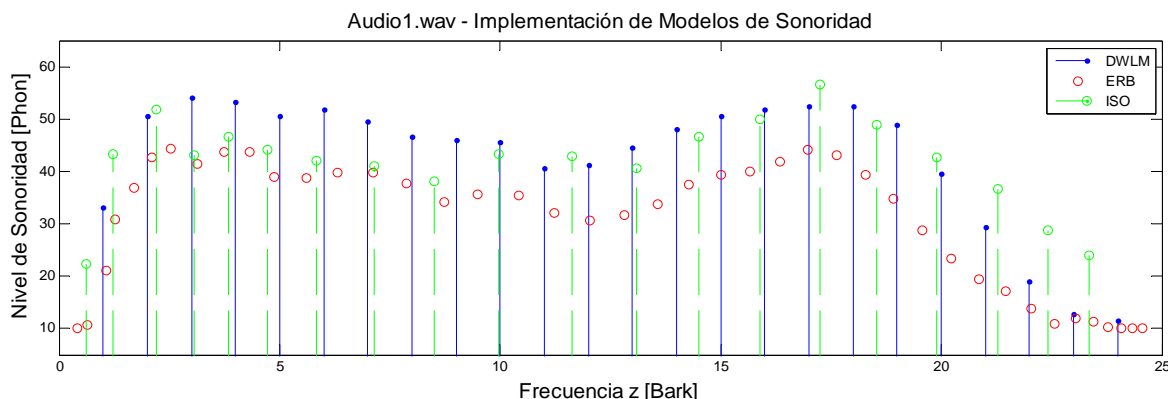


Figura 6. En azul se presenta el modelo propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Los valores de sonoridad S y nivel de sonoridad L_s que se obtienen para cada modelo son, respectivamente: $S = 33,1$ Sones y $L_s = 90,5$ Fones; $S = 26,0$ Sones y $L_s = 87,0$ Fones; $S = 22,6$ Sones y $L_s = 84,9$ Fones.

En las figuras (6) y (7) es posible notar que el modelo coclear de Patterson-Holdsworth muestra una sonoridad por banda menor que los otros modelos debido a que el rango auditivo lo divide en una mayor cantidad de bandas. Pese a ello, los tres modelos muestran similares resultados respecto de las regiones de frecuencias que son subjetivamente más audibles. En *Audio1.wav*, estas regiones se encuentran entre 3 y 4 Barks (300 a 400 Hz) y 17 y 18 Barks (en torno a 4 kHz). Para *Audio2.wav*, las regiones están entre 6 y 7 Barks (entre 600 y 900 Hz) y entre 16 y 17 Barks (en torno a 3,5 kHz).

El modelo propuesto entrega una mayor ponderación a las bajas frecuencias que los otros 2 modelos, esencialmente en el rango de frecuencias de 3 a 7 Barks (300 a 900 Hz). Estas diferencias pueden atribuirse a que el ancho de banda usando la técnica de *warping* es mayor que los valores experimentales publicados en (Zwicker, 1961). Por otra parte, el modelo propuesto considera los contornos de igual sonoridad actualmente vigentes.

⁴ Para el cálculo de sonoridad por cada banda crítica se usó la ecuación (11), pero considerando como límites de integración las frecuencias de corte de cada una de las 20 bandas críticas aproximadas en el Método B.

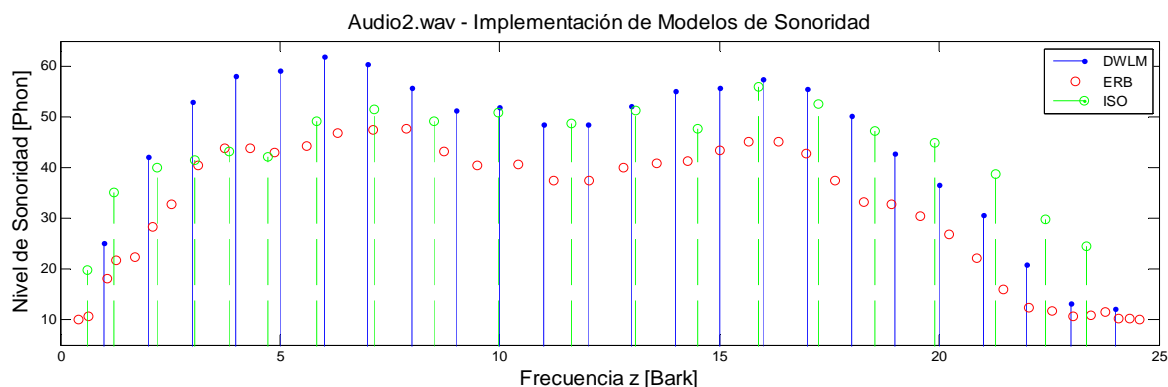


Figura 7. En azul se presenta el modelo propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Los valores de sonoridad S y nivel de sonoridad L_s que se obtienen para cada modelo son, respectivamente: $S = 41,1$ Sones y $L_s = 93,6$ Fones; $S = 29,0$ Sones y $L_s = 88,6$ Fones; $S = 25,1$ Sones y $L_s = 86,5$ Fones.

6 Conclusiones

El presente trabajo ha mostrado que es posible obtener la sonoridad de una señal de audio en condición de campo libre utilizando una metodología basada en Redes Neuronales Artificiales (RNA) para la ponderación de sonoridad y en la técnica de *warping* de frecuencias para la aproximación a las bandas críticas del oído humano, entregando valores de sonoridad análogos a los que se obtienen usando el Método B del estándar de referencia ISO 532-B. La RNA fue entrenada e implementada considerando el estándar ISO 226:2003, por lo que actuará como interpoladora de los contornos de igual sonoridad actuales, en contraste de lo que ocurre con el estándar ISO 532 que, estando vigente, no considera esta actualización.

Referencias

- Bladon, R.A.W. (1981). "Modeling the judgment of vowel quality differences". Journal of the Acoustical Society of America, Vol. 69(5), pp 1414-1422.
- Espinoza, V.; Venegas, R.; Floody, S. (2006). "Modelo de sonoridad usando redes neuronales artificiales". V Congreso Iberoamericano de Acústica, FIA, Santiago, Chile.
- Espinoza, V. (2008). "Estimador de distorsión no lineal usando características psicoacústicas aplicado a una situación de altavoces". VI Congreso Iberoamericano de Acústica, FIA, Buenos Aires, Argentina.
- ISO 226:2003 Acoustics – Normal Equal-loudness-level contours.
- ISO 532:1975 Acoustics – Method for calculating loudness level.
- Osses, A.; Espinoza, V. (2009). "Análisis espectral en bandas críticas usando la técnica de *Warping* de frecuencias". Anais do 7º Congresso da AES Brasil, pp 73-76.
- Patterson, R.D; Robinson, K.; Holdsworth, J.; McKeown, D; Zhang, C; Allerhand, M.(1992). "Complex sounds and auditory images". In Auditory physiology and Perception, Proc. 9th International Symposium on Hearing. (Editors) Cazals, Y., Demany, K., Horner, K., Pergamon, Oxford.
- Smith III, J.; Abel, J. (1999). "Bark and ERB bilinear transforms". IEEE Tr. Speech and Audio Processing, pp 697-708.

- Suzuki, Y.; Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *Journal of the Acoustical Society of America* 116, pp 918-933.
- Terhardt, E.; Stoll, G.; Seewann, M. (1982). "Pitch of complex signals according to virtual-pitch theory: tests, examples and predictions". *Journal of the Acoustical Society of America*, Vol. 71(3), pp 679-688.
- Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands". *Journal of the Acoustical Society of America*, Vol. 33(2), pp 248.
- Zwicker, E.; Terhardt, E. (1980). "Analytical expressions for critical band rate and critical bandwidth as a function of frequency". *Journal of the Acoustical Society of America*, Vol. 68(6), pp 1523-1525.
- Zwicker, E.; Fastl, H.; Widmann, U.; Hurakata, K.; Kuwano, S.; Namba, S. (1991). "Program for calculating loudness according to DIN 45631 (ISO 532B)". *Journal of the Acoustical Society of Japan*, Vol. 12(1), pp 39-42.
- Zwicker, E.; Fastl, H. (2007). "Psychoacoustics, facts and models". 3rd Ed., Springer-Verlag, Berlín, Alemania.