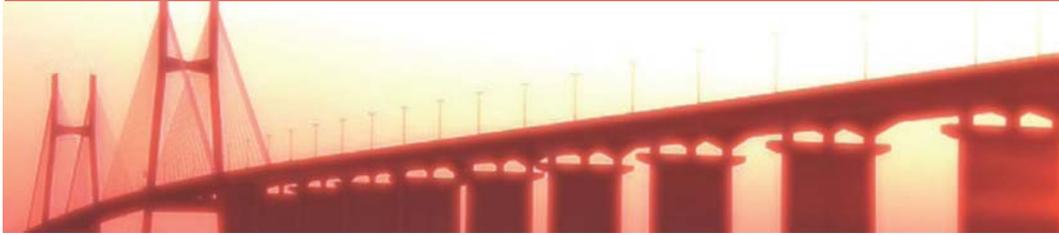


# PRIMERAS JORNADAS REGIONALES DE ACÚSTICA AdAA 2009

19 y 20 de noviembre de 2009, Rosario, Argentina



AdAA2009-A027R

## **Ponderación psicoacústica de la distorsión no lineal de dispositivos de audio: una propuesta para su medición**

Víctor Espinoza Catalán <sup>(a), (b)</sup>, Alejandro Osses Vecchi <sup>(b)</sup>.

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

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

### **Abstract**

This paper researches the possibilities of using new techniques for the measurement of audio devices. In particular, the use of frequency sweeps offer the main advantages of giving a large dynamic range and the possibility to separate the linear and the non-linear responses of the system under test (SUT). By means of this separation of responses, it is possible to implement a non-linear distortion estimator considering the psychoacoustics characteristics of loudness and critical bands. A loudness model based on Artificial Neural Networks (ANN) will be used, as well as a simultaneous masking model and an ERB critical-band spectral integration. Some examples of the measurement of audio devices and electrodynamic loudspeakers are presented.

### **Resumen**

Este trabajo investiga las posibilidades de uso de nuevas técnicas de medición para dispositivos de audio. En particular el uso de barridos de frecuencia, que ofrecen como principal ventaja un gran rango dinámico y la posibilidad de separar la respuesta lineal de la respuesta no lineal del sistema bajo prueba (SBP). Mediante esta separación de respuestas, es posible implementar un estimador de distorsión no lineal usando características psicoacústicas de sonoridad y bandas críticas. Se utiliza un modelo de sonoridad basado en Redes Neuronales Artificiales (RNA), un modelo de enmascaramiento simultáneo y una integración espectral en bandas críticas ERB. Se presentan ejemplos de medición para dispositivos de audio y altavoces electrodinámicos.

## 1 Introducción

Hasta ahora existen múltiples métodos para poder determinar la influencia en la calidad de dispositivos de audio frente a la distorsión armónica. Si bien el “Total Harmonic Distortion (THD)” considera solo la relación de energía de entrada y salida que pasa por un sistema bajo prueba (SBP), sin ninguna ponderación auditiva sobre sus datos, este continúa siendo la estimación común por muchos fabricantes en la industria del audio. En una aplicación específica para altavoces, Temme (2000) propone el %Rub&Buzz, que utiliza un concepto similar al THD pero sobre un rango de frecuencias con mayor probabilidad de escucha. Otras técnicas de medición más elaboradas consideran el uso de múltiples tonos (Temme 2008) en relación no armónica para evidenciar no solo las componentes de distorsión armónicas sino también las intermoduladas. Si bien esta última es una señal más cercana a una señal compleja, los datos rescatados de esta prueba no consideran el peso de la componente auditiva.

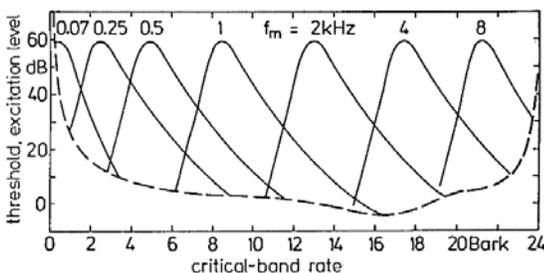
Por lo anterior, este trabajo pretende aportar sobre esta problemática de mediciones de sistemas electroacústicos, con la incorporación de otros factores para el análisis de la información. La propuesta consiste en eliminar y ponderar la información psicoacústica no relevante a través de modelos de percepción auditiva, que en conjunto con una técnica de medición de respuesta impulso, hace posible obtener de manera rápida y significativa, datos discriminados con los modelos anteriormente mencionados.

En particular se utilizarán tres modelos de características psicoacústicas del oído: enmascaramiento simultáneo por tonos, escalamiento en Barks, y nivel de sonoridad ponderado en ERB. La técnica de medición para el SBP consiste en la creación de barridos de frecuencia logarítmicos que permiten separar la componente lineal de la no lineal (distorsión armónica) por medio de la aplicación de un filtro inverso del barrido inicial. La implementación se realizará con el Data Acquisition Toolbox, el Signal Processing Toolbox, y el Neural Network Toolbox de MATLAB v7.6.0.

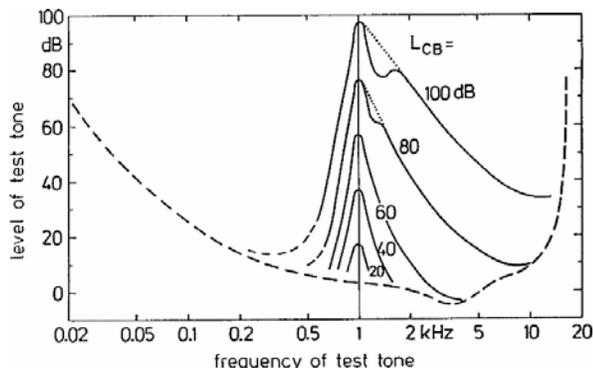
## 2 Antecedentes Psicoacústicos

### 2.1 Enmascaramiento simultáneo

En figura 1 y figura 2 se grafica el enmascaramiento producido por bandas críticas en escala Bark. Se observa que las pendientes de las curvas para cada banda crítica son similares y la pendiente superior es variable con el nivel de excitación.



**Figura 1.** Curvas de enmascaramiento para varios Bark (Zwicker y Fastl, 2007).



**Figura 2.** Variación de la banda crítica centrada en 1 kHz a distintos niveles.

Visto esto, Terhardt (1979) propone el uso de un modelo simple con rectas de pendientes variables con el nivel y escala Bark en ecuación 1.

$$\begin{aligned}
 S_1 &= 27 \text{ dB/Bark} \quad ; \quad f_M > f_T \\
 S_2 &= -(24 + 230/f - 0,2 \times L) \text{ dB/Bark} \quad ; \quad f_M \leq f_T
 \end{aligned}
 \tag{1}$$

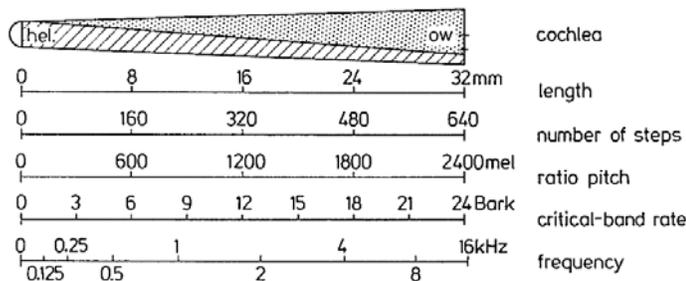
donde

- $f$  : frecuencia en Hz
- $f_M$  : frecuencia tono enmascarador en Hz.
- $f_T$  : frecuencia en banda crítica en Hz.
- $L$  : nivel de tono enmascarador en dB NPS

Para la escala relativa a la altura (pitch-related scale) se utilizará la estimación de Zwicker y Terhardt (1980), que está en directa relación con la posición de excitación de la membrana basilar del oído (ver figura 3 y ecuación 2).

$$z = 13 \times \text{atan}(0,76 \times f) + 3,5 \times \text{atan}(f / 7,5)^2
 \tag{2}$$

- $f$  : frecuencia en kHz.
- $z$  : escala Bark.



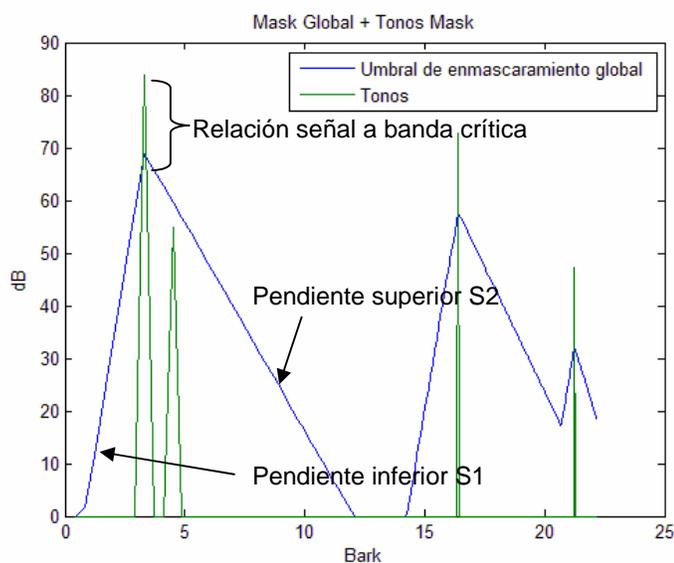
**Figura 3.** Membrana basilar con medidas más usadas (Zwicker y Fastl, 2007).

Para el enmascaramiento global, y sin considerar la adición de bandas críticas, se obtiene,

$$Mask_{Global} = \max(Mask(f_T)), \tag{3}$$

siendo  $Mask(f_T)$  la curva de enmascaramiento para cada frecuencia de prueba.

Debido a que las componentes armónicas se encuentran sobre la frecuencia de prueba, solo la pendiente superior (S2) del modelo de Terhardt estará implementada. Para la relación señal a banda crítica en figura 4, Johnston (1988) presenta una aproximación de 5.5 dB para enmascaramiento tonal.



**Figura 4.** Modelo de enmascaramiento. Se presentan rectas de enmascaramiento para 3 tonos. Notar que un tono está bajo la curva global y no genera enmascaramiento

## 2.2 Modelo de Nivel de Sonoridad

Se empleará el modelo de sonoridad desarrollado por Espinoza et al (2006). El modelo consiste en una estimación de nivel de sonoridad para cualquier frecuencia y nivel de presión sonora, entre 0-90 fones y 20-12500 Hz, mediante una interpolación usando redes neuronales artificiales (RNA). Los datos de entrenamiento (curvas isofónicas) fueron los publicados por Suzuki y Takeshima (2004) los cuales actualizaron la norma ISO 226:2003. La configuración actual de la RNA es de 19 neuronas en la capa de entrada, 20 en la capa oculta y una de salida. El método de entrenamiento fue feedforward Quasi-Newton Backpropagation, con error objetivo de  $10^{-5}$ . El error en banda ancha es de  $\pm 1$  dB sobre 100 Hz y  $\pm 3$  dB bajo 100 Hz. Posterior a la aplicación del modelo se transforman el nivel de sonoridad a sonoridad, mediante la aproximación,

$$S = \begin{cases} 2^{\frac{(P-40)}{10}} & \text{Si } P \geq 40 \\ \left(\frac{P}{40}\right)^{2,642} & \text{Si } P \leq 40 \end{cases} \quad (4)$$

siendo P el nivel de sonoridad a 1 kHz, y S la sonoridad en sones. (Bladon 1981).

### 2.3 Equivalent Rectangular Bandwidth (ERB)

Consiste en el área equivalente de un filtro rectangular a un filtro perceptual de banda crítica el cual aproxima a 41 bandas de filtro rectangular todo el espectro de frecuencias audibles (Moore y Glasberg 1996). Se seguirá un método similar al empleado en ISO 532:1975-B para la suma de sonoridad específica por banda ERB. El uso de ERB se emplea en la integración de sonoridad específica de los armónicos audibles por banda crítica para su graficación en nivel de sonoridad por banda. Para poder transformar las cantidades antes mencionadas se utilizarán las siguientes ecuaciones:

$$ERB = 0,108 \times f_c + 24,7 \quad (5)$$

$$ERB_N = 21,4 \times \log_{10}(0,0047 \times f + 1) \quad (6)$$

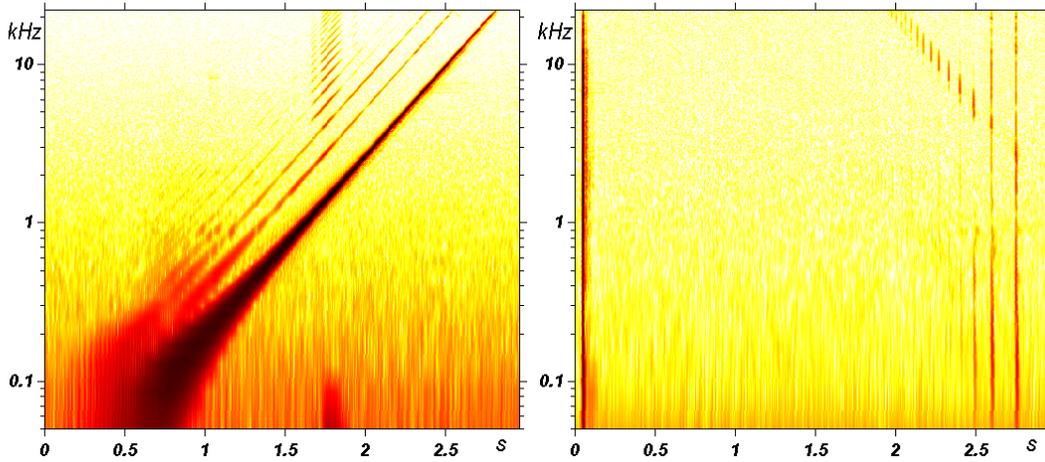
$f_c$  : frecuencia central banda ERB en Hz, asociado al valor central de la banda

$f$  : frecuencia en Hz

$ERB_N$  : Número  $ERB$ , asociado a la banda ERB

### 2.4 Medición de función de transferencia usando Sweep

Farina (2000) desarrolla la posibilidad de separar las componentes lineal y no lineal de la respuesta impulso por medio del análisis en series de Volterra del SBP. Massarani y Muller explican las múltiples maneras y aplicaciones derivadas del uso de barridos de frecuencia para medición de función de transferencia y síntesis de barridos en dominio de la frecuencia. El método consiste en generar un barrido de frecuencia logarítmico de espectro de frecuencia de -3dB por octava. Esta señal permite establecer una separación de tiempo constante de las componentes armónica con respecto a la respuesta lineal. Con la aplicación del filtro inverso del barrido es posible separar cada componente de distorsión armónica en tiempos negativos a la respuesta impulsiva lineal. Un ejemplo se grafica en la figura 5.



**Figura 5.** Espectrograma de barrido de frecuencia. Izquierda: mostrando distorsión armónica. Derecha: luego de aplicar filtro inverso. En este último y cercano a 0 y 2.5 segundos se aprecia la respuesta impulso lineal y la respuesta impulso de los armónicos, respectivamente. (Massarani y Muller).

### 3 Implementación

Sean  $x[n]$ ,  $y[n]$  secuencias de entrada y salida del SBP, donde  $n = 0 \leq n \leq N-1$ . La secuencia  $x[n]$  debe ser un barrido de frecuencia logarítmico para poder separar posteriormente las respuestas impulso de la distorsión armónica.

Para evitar el posible solapamiento temporal usando convolución en el dominio de la frecuencia es necesario transformar las señales de  $x[n]$ ,  $y[n]$  en  $x_p[n]$ ,  $y_p[n]$  como:

$$x_p[n] = \begin{cases} x[n], & 0 \leq n \leq N-1 \\ 0, & N \leq n \leq 2N-1 \end{cases} \quad (7)$$

$$y_p[n] = \begin{cases} y[n], & 0 \leq n \leq N-1 \\ 0, & N \leq n \leq 2N-1 \end{cases} \quad (8)$$

Esto es, agregar  $N$  ceros a las secuencias  $x[n]$ ,  $y[n]$ . Lo anterior es necesario para calcular un equivalente a una convolución lineal para las primeras  $N$  muestras de la señal. Detalles en Oppenheim et al. (2000).

Para determinar la función de transferencia y la respuesta impulso del SBP se utiliza la transformada rápida de Fourier. Para la función de transferencia,

$$H[k] = \frac{FFT(y_p[n])}{FFT(x_p[n])}, \quad (9)$$

y para la respuesta impulso,

$$h[n] = IFFT(H[k]). \quad (10)$$

Truncando esta respuesta impulso  $h[n]$  a las  $N$  primeras muestras obtenemos la respuesta impulso lineal del sistema. En las restantes muestras se puede obtener las respuestas impulso de la distorsión armónica.

Determinada la respuesta  $h[n]$  es necesario separar los diferentes impulsos para su análisis. Si el retardo producido por las condiciones de medición es compensado, la posición de cada armónico esta dada por (Farina 2000):

$$n_{armónico} = N - \frac{N/2}{fs} \times \frac{\log_2(n_a)}{s_r}, \quad (11)$$

donde:

$$s_r = \log_2\left(\frac{f_2}{f_1}\right) \quad ; \quad \text{tasa de cambio barrido (sweep rate)}$$

$f_2$  = frecuencia final barrido

$f_1$  = frecuencia inicial barrido

$n_a$  = número de armónico  $\geq 2$

Además la duración de cada armónico es:

$$\Delta n_{armónico} = \frac{N/2}{fs} \times \frac{\log_2\left(\frac{n_a}{n_a-1}\right)}{s_r}. \quad (12)$$

Como es observado, las respuestas impulso armónicas tienen distinta cantidad de muestras. Para hacerlas comparables se agregan ceros hasta completar  $N$  muestras.

Finalmente, para graficar la respuesta de frecuencia de cada impulso se utilizará:

$$dB = 10 \cdot \log_{10}\left(|H_i[k]|^2\right) \quad \begin{array}{l} k = 0, 1, 2, \dots, (N/2 - 1) \\ i = 1, 2, \dots, n_{am} \end{array}$$

donde:

$n_{am}$  = número de armónicos a medir

Para las mediciones a presentar, se usó un barrido de frecuencia de 0.1 Hz a 24000 Hz con 65536 muestras. Donde fue necesario se aplicó una ventana Tukey asimétrica, evitando pérdidas en el espectro de frecuencias de cada impulso.

Posterior a esto se aplica la curva de enmascaramiento global generada por la respuesta de frecuencia lineal a cada respuesta de frecuencia armónica. A continuación, se calcula el nivel de sonoridad con el modelo RNA usando una curva suavizada de datos a 48° de octava

para reducir el exceso de cálculo y detalle, seguido de la transformación de nivel de sonoridad a sonoridad integrando por banda ERB y graficando en nivel de sonoridad nuevamente.

### 4 Resultados

Se realizaron pruebas con sistemas de altavoces amplificados, procesador de dinámica digital y altavoz al aire libre. El equipamiento utilizado fue un micrófono Beyerdynamic MM1, tarjeta de sonido Presonus Audiovox y amplificador Alesis RA-100, en cámara anecoica. Se presentan las respuestas de frecuencia lineal y no lineal, sin y con ponderación.

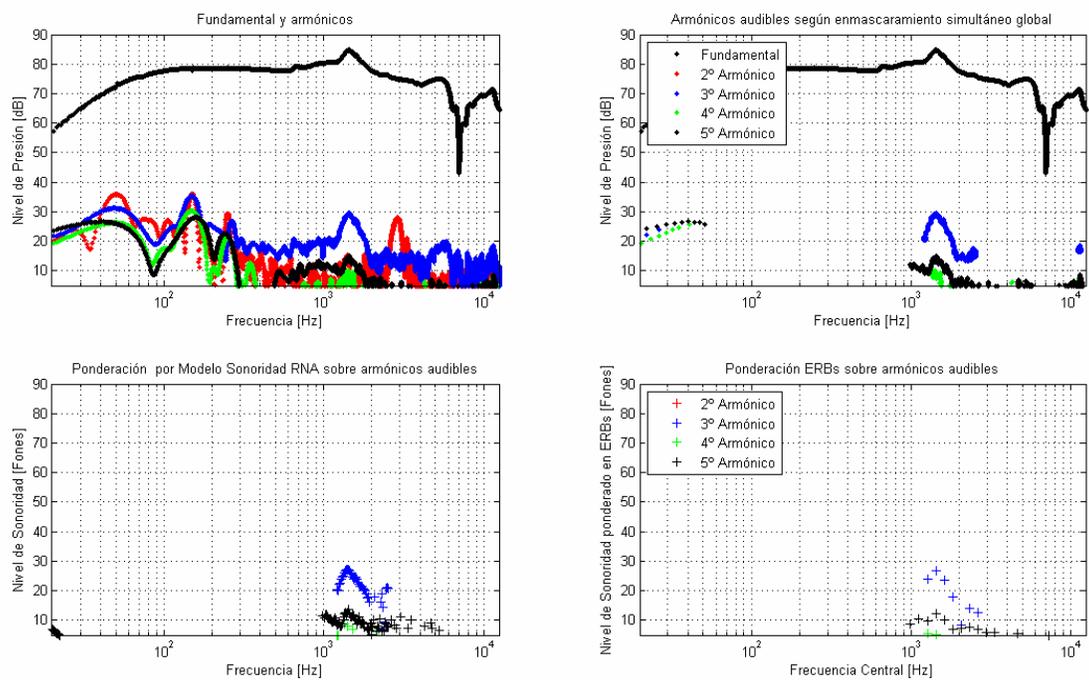


Figura 6. Medición altavoz aire libre

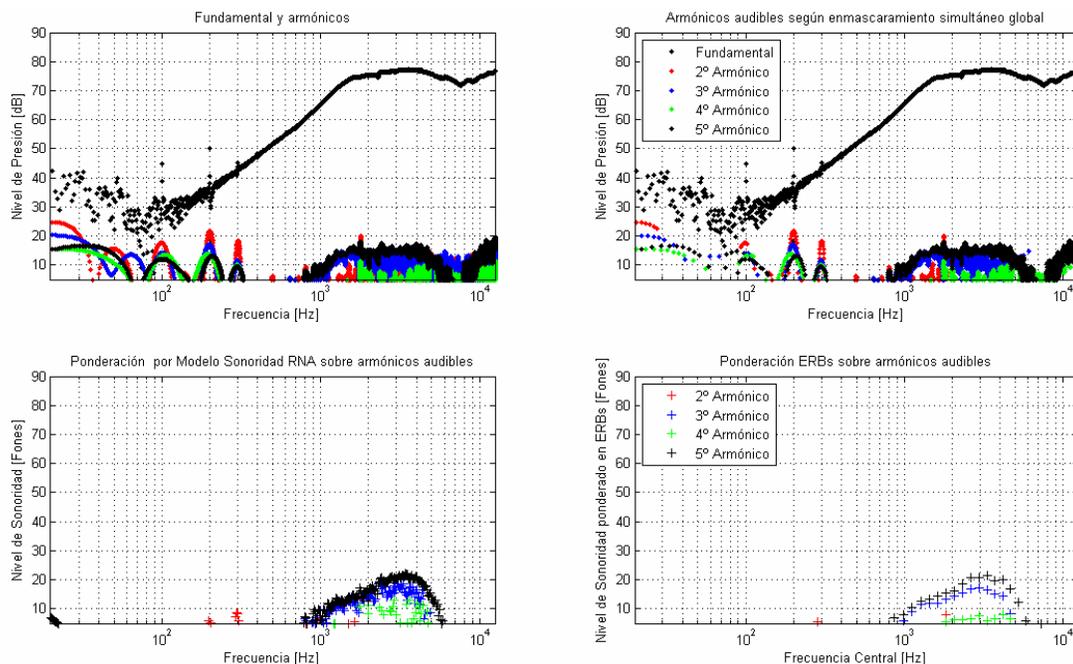


Figura 7. Tweeter en campo cercano

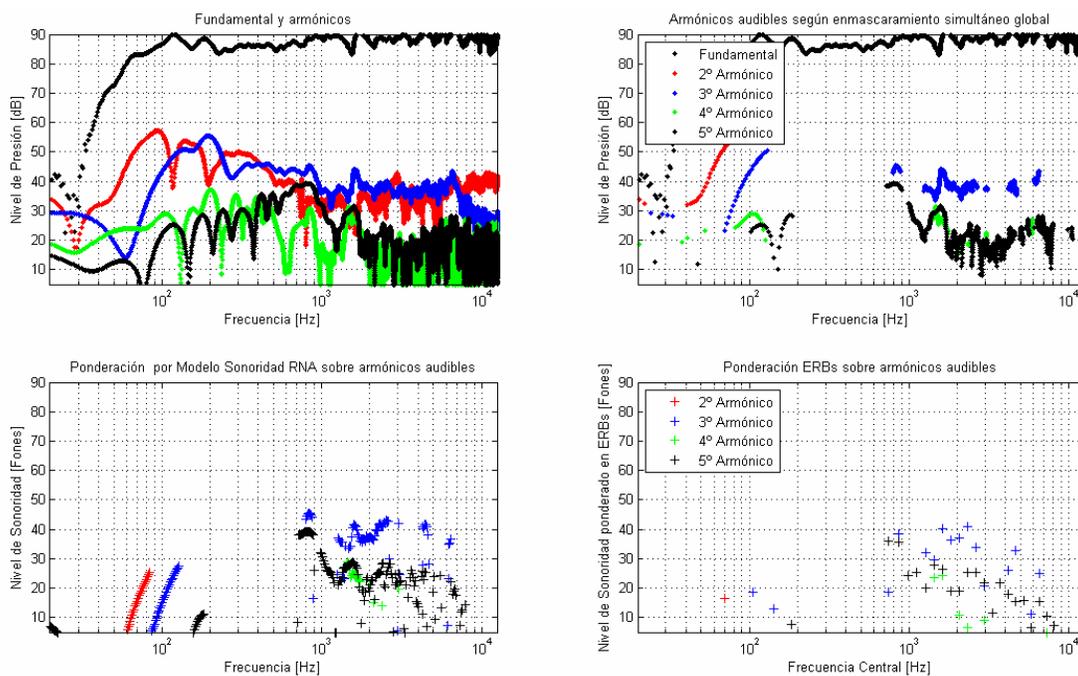
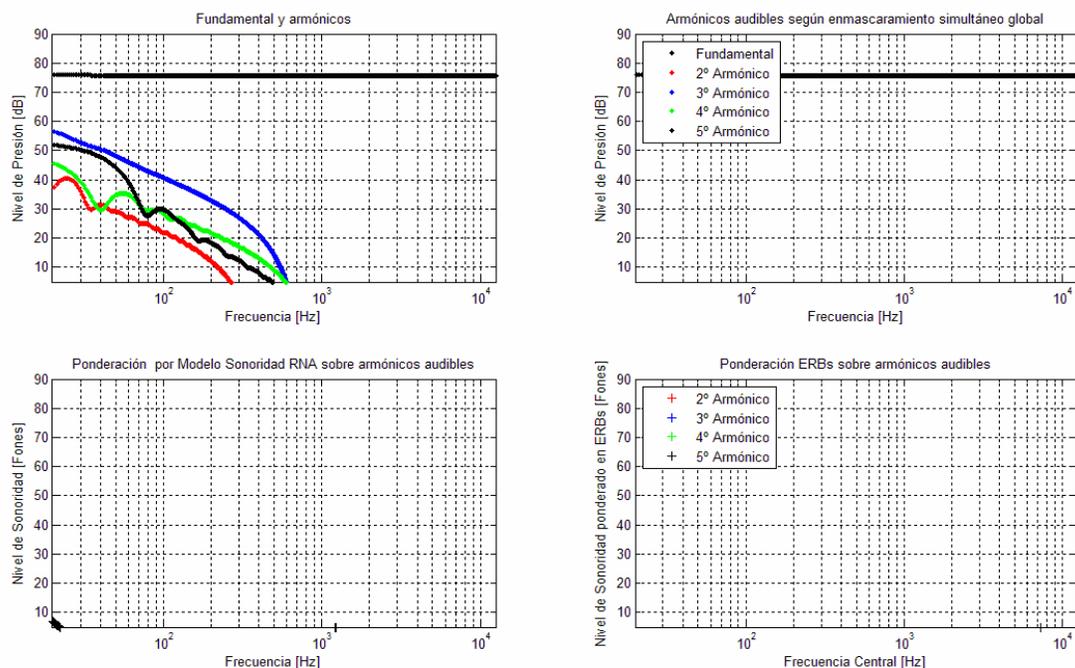


Figura 8. Medición monitor estudio bi-amplificado



**Figura 9.** Medición procesador de dinámica digital

## 5 Conclusiones

Se ha presentado un procedimiento para medir distorsión audible utilizando modelos psicoacústicos conocidos. De las mediciones es observado que los mayores niveles de distorsión no necesariamente tienen influencia audible analizando la zona de baja frecuencia.

Al ser eliminadas las componentes de frecuencia no audible, el análisis de resultados es más claro. Esto podría discriminar que sistema tiene mayor fidelidad en la reproducción de una señal, entregando información útil para orientar en algunas decisiones de diseño (por ejemplo, frecuencia de corte crossover), en la necesidad de establecer un umbral para el control de calidad de determinado producto en la industria del audio, o en la elección de componentes.

En trabajos futuros, debiera incorporarse la distorsión por intermodulación y características auditivas más elaboradas como diferencias tonales y no tonales en el enmascaramiento y así también modelos de codificación neuronal de la información auditiva (Zwicker y Fastl, 2007).

## 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.
- Farina, A. (2000); "Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine technique", *J.AES*, vol. 48, p. 350, 108th AES Convention, Paris, Preprint 5093.

- ISO 226:2003 Acoustics – Normal Equal-loudness-level contours.
- ISO 532:1975 Acoustics – Method for calculating loudness level.
- Johnston, J. (1988). Transform Coding of Audio Signals Using Perceptual Noise Criteria, IEEE J. Selected Areas in Communications, Vol. 6, No. 2, pp. 314–323.
- Massarani, P.; Muller, S. “Transfer-Function Measurement with Sweeps: Director’s cut including previously unreleased material”.
- Metzler, Bob (2005) “Audio Measurement Handbook”, Audio Precision, Oregon, USA.
- Moore, B.; Glasberg, B. (1996) “A revision of Zwicker's loudness model,” *Acta Acustica*, vol. 82, pp. 335-345.
- Oppenheim A.;Schafer, R.; Buck, J. (2000). “Tratamiento de señales en tiempo discreto”. 2ª Ed., Pearson Education, pp 579-581.
- 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.
- Suzuki, Y.; Takeshima, H. (2004). Equal-loudness-level contours for pure tones. Journal of the Acoustical Society of America 116, pp 918-933.
- Temme, Steve (2000); “Are You Shipping Defective Loudspeakers to Your Customers? ”, Presented at the ALMA 2000 Winter Symposium, Las Vegas, Nevada, USA.
- Temme, S.; Brunet, P.; 2008. “A New Method for Measuring Distortion Using a Multitone Stimulus and Noncoherence” JAES Volume 56 Issue 3 pp. 176-188.
- Terhardt, E. (1979). “Calculating virtual pitch”. Hearing Research, 1:155–182.
- 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. (2007). “Psychoacoustics, facts and models”. 3<sup>rd</sup> Ed., Springer-Verlag, Berlín, Alemania.