

## MODELO PARA LA PREDICCIÓN DE LA ESTRUCTURA TEMPORAL Y ESPECTRAL DE LOS PAISAJES SONOROS URBANOS BASADO EN REDES NEURONALES

PACS: 43.50.Rq

Torija, Antonio J.<sup>1</sup>; Ruiz, Diego P.<sup>1</sup>; Ramos-Ridao, Ángel<sup>2</sup>

<sup>1</sup>Dpto. Física Aplicada Facultad de Ciencias

<sup>2</sup>Dpto. Ingeniería Civil E.T.S. de I.C.C.P.

Campus Fuentenueva s/n

Universidad de Granada

18071 Granada

Tel: 958 240 771

Fax: 958 243 214

E-mail: [ajtorija@ugr.es](mailto:ajtorija@ugr.es); [druiz@ugr.es](mailto:druiz@ugr.es); [ramosr@ugr.es](mailto:ramosr@ugr.es)

### ABSTRACT

The urban soundscapes are dominated by the noise generated by the road traffic, which acts as a noisy source that masks the rest of present sounds, which creates diverse problems that go beyond that the annoyance and displeasure in the population. On the other hand, the perception of the soundscapes by the population is not explained only for the sound pressure level to which it is exposed, but temporal and spectral characteristics of the soundscape must be taking into account, as well as other variables not acoustics, e.g., the attitude towards the source, sensitivity and habituation to the noise, meant to the population, etc. All this supposes that, as different authors establish and, within the framework of the current European legislation, the design, planning and assessment of the soundscapes should be included in the future town planning. In this paper a model, based on neural networks, for the prediction of the temporal and spectral structure of the soundscapes is proposed, which, together with the perceptual assessment of the soundscapes, allows us the adjustment of these to the needs and requirements of the population in the urban environments. Likewise, a comparison between the predicted and the experimentally obtained data in diverse sound scenes of the city of Granada is presented.

### RESUMEN

Los paisajes sonoros urbanos están dominados por el ruido generado por el tráfico rodado, que actúa como una fuente ruidosa que enmascara el resto de sonidos presentes, lo cual crea diversos problemas que van más allá de la molestia y desagrado en la población. Por otro lado, la percepción que la población tiene de los paisajes sonoros no se explica solamente por el nivel de presión sonora a que está expuesta, sino que entran en juego las características temporales y espectrales del paisaje sonoro, así como otras variables no acústicas como pueden ser, la actitud hacia la fuente, sensibilidad y habituación al ruido, significado para la población, etc. Todo ello supone que, tal y como establecen diferentes autores y, dentro del marco de la actual legislación Europea, el diseño, planificación y evaluación de los paisajes

sonoros debería ser incluido en la futura planificación urbana. En esta comunicación se propone un modelo, basado en redes neuronales, para la predicción de la estructura temporal y espectral de los paisajes sonoros, el cual, junto con la evaluación perceptual de los paisajes sonoros, nos permita la adaptación de estos a las necesidades y requerimientos de la población en los entornos urbanos. Se presentan asimismo una comparación entre los datos predichos y los obtenidos experimentalmente en diversos escenarios sonoros de la ciudad de Granada.

## **1. INTRODUCCIÓN**

Tradicionalmente, la evaluación del ruido ambiental se ha centrado en la reducción del nivel de ruido de sonidos indeseados. Más recientemente, con la introducción del concepto de paisajes sonoros [1], el problema se enfoca desde un punto de vista más holístico, de manera que el ambiente acústico urbano es observado como una multitud de sonidos los cuales pueden evocar emociones específicas, siendo considerado, por lo tanto, el paisaje sonoro urbano como una parte integral del ambiente vital urbano [2]. La evaluación y descripción de un paisaje sonoro debería, por tanto, tener en cuenta la relación entre la experiencia individual y subjetividad, con el contexto socio-cultural y físico en el que se desarrolla. Es en este último punto en el que se centra este trabajo.

La gran complejidad de las aglomeraciones urbanas supone un problema difícil de solventar desde el punto de vista de la predicción de las variables acústicas ambientales. En los entornos urbanos aparecen una gran variedad de fuentes sonoras y espacios, los cuales generan una gran heterogeneidad de situaciones acústicamente relevantes. Así, localizaciones con diferente composición de tráfico rodado, diferentes configuraciones urbanas, presencia de fuentes ruidosas difíciles de caracterizar (por ejemplo, actividades comerciales o de ocio, obras de construcción,...), existencia de espacios verdes, etc., genera la aparición de espacios sonoros con diferentes niveles de presión sonora, así como con grandes diferencias en cuanto a su composición temporal y espectral. La heterogénea fisonomía de los espacios urbanos junto con las características del ruido ambiental, con una gran variabilidad espacial, temporal y espectral, hace de su modelización y predicción un problema complejo y no-lineal, lo cual obliga a la utilización de una potente herramienta de minería de datos, las redes neuronales artificiales [3]. Las redes neuronales artificiales son un paradigma de aprendizaje y procesamiento automático que busca como último objetivo emular el comportamiento cerebral, o al menos algunas de sus funciones, tales como el aprendizaje. La utilización de este método proporciona flexibilidad, exactitud y una gran capacidad de adaptación, lo cual resulta muy interesante a la hora de abordar la modelización de las características acústicas de un entorno urbano.

Por esta razón, en esta investigación hemos desarrollado un modelo, basado en redes neuronales de retro-propagación, para obtener de una adecuada predicción de las características temporales y espectrales, así como de los niveles sonoros, para las diferentes localizaciones urbanas consideradas. El objetivo de este modelo es proporcionar una herramienta para el diseño, planificación y evaluación de los paisajes sonoros urbanos con la meta final de incorporar las necesidades de la población y obtener aglomeraciones urbanas acústicamente sostenibles.

## **2. METODOLOGÍA**

### **2.1 Proceso de obtención de datos**

Para el estudio de los distintos paisajes sonoros de la ciudad de Granada, planteamos la idea de crear un sistema que, a través de redes neuronales, nos ayudara a la obtención de resultados precisos. Fueron seleccionadas 80 localizaciones de la ciudad de Granada, tratando incluir la gran heterogeneidad de situaciones acústicamente relevantes característica de los

entornos urbanos. Situaciones con diferentes condiciones de tráfico rodado, en distintos periodos del día, con diferente geometría y fisonomía de la localización, en parques y plazas urbanas (sin la presencia de tráfico rodado), con la presencia de fuentes ruidosas diferentes del tráfico rodado (obras de construcción, actividades comerciales,...), etc., fueron seleccionadas. Una vez realizadas todas las mediciones, cada 5 minutos, se llevó a cabo el cálculo de los diferentes indicadores acústicos, de manera que, a partir de las diferentes variables de entrada se realizó la predicción de cada uno de los descriptores acústicos utilizados, con 5 minutos como periodo de integración.

## 2.2 Variables de entrada y salida del modelo

Para el desarrollo del modelo de predicción, como podemos observar en la tabla 1, han sido seleccionadas 33 variables de entrada, incluidas en 2 módulos, un módulo de propagación sonora y otro de emisión sonora, el cual, a su vez, está formado por 3 sub-módulos. Estos sub-módulos contienen variables para la caracterización del entorno, variables relacionadas con la evolución temporal del nivel sonoro y variables para la caracterización de la emisión sonora generada por el tráfico rodado, respectivamente. Además, tenemos 36 variables de salida, el nivel de presión sonora ponderado A ( $L_{Aeq}$ ), el nivel de presión sonora no ponderado ( $L_{eq}$ ), la diferencia entre ambos descriptores, la variabilidad temporal y la impulsividad del nivel sonoro, así como el nivel sonoro para cada una de las bandas de 1/3 de octava entre 20 Hz – 20 KHz.

Tabla 1. Variables de entrada y salida del modelo de predicción. \*Entre paréntesis se muestran los rangos de valores de las distintas variables de salida.

Variables de entrada			Variables de salida
Módulo	Sub-módulo	Variable	
Emisión sonora	Entorno	Tipo de día	
		Periodo del día	
		Actividad comercial/ocio	
		Aparición de obras de construcción	
		Tipo de localización	
	Evolución temporal del nivel sonoro	Aparición de fuentes de agua	
		Presencia de vegetación	
		Tiempo de estabilización	$L_{Aeq}$ (0-100 dBA)
		Tipo de flujo de tráfico	$L_{eq}$ (0-110 dB)
		Eventos sonoros anómalos relacionados con el tráfico	$L_{eq} - L_{Aeq}$ (0-30)
	Emisión sonora del tráfico rodado	Eventos sonoros anómalos no relacionados con el tráfico	TSLV (0-60 dB <sup>2</sup> )
		Caudal de tráfico rodado ascendente/descendente	CF (0-10)
		Número de vehículos con sirena	
		Velocidad media	
		Pendiente de circulación	Bandas de 1/3 octava entre 20-20000 Hz (0-90 dB)
Propagación sonora	Número de carriles ascendentes/descendentes		
	Tipo de pavimento		
	Estado de la superficie de rodadura		
	Geometría de la calle		
	Anchura de la vía		
	Altura de la vía		
	Anchura de la calzada		
Distancia fuente-receptor			

Con respecto al caudal de tráfico rodado, han sido considerados 5 tipos de vehículos: vehículos ligeros, vehículos pesados, autobuses, motocicletas-ciclomotores y vehículos de saneamiento urbano. Para la evaluación de la variabilidad temporal y de la impulsividad del nivel de presión sonora, en este trabajo han sido utilizados los descriptores Varianza Temporal del Nivel Sonoro (TSLV) y Factor Cresta (CF), respectivamente [4].

### 2.3 Estructura del modelo basado en redes neuronales de retro-propagación

Por sus propiedades para generalizar y su posibilidad de realizar un aprendizaje supervisado, hemos utilizado una red neuronal artificial, con la variante Levenberg-Marquardt con retro-propagación con regulación Bayesiana (Función de entrenamiento Trainbr). La estructura de esta ANN (mostrada en la figura 1) es 33 variables de entrada, 36 neuronas en la capa oculta y 36 variables de salida.

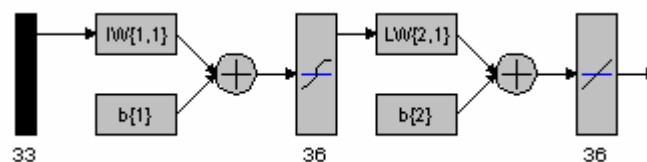


Figura 1. Estructura de la red neuronal propuesta.

Una vez seleccionada la estructura de la red neuronal, mostrada en la figura 1, construimos 5 conjuntos de datos diferentes. Los 5 conjuntos de datos fueron creados aleatoriamente a partir de los 533 registros de entrada. Por lo tanto, obtuvimos 5 conjuntos de datos, cada uno de los cuales contenía un conjunto de entrenamiento y un conjunto de validación, formados por registros diferentes. Los conjuntos de entrenamiento y validación contienen 350 y 183 registros, respectivamente. La ANN fue ejecutada 5 veces, cada una con diferentes pesos iniciales, con cada uno de los 5 conjuntos de datos. Esto nos ofrece 25 pruebas para evaluar la precisión de la predicción de los diferentes descriptores acústicos utilizados en este trabajo.

## 3. RESULTADOS

### 3.1 Análisis de la relación entre las variables de entrada y los descriptores acústicos utilizados

Tabla 2. Variables de entrada con mayor nivel de correlación con los descriptores acústicos utilizados.

\*Entre paréntesis se especifica el nivel de correlación mediante el factor  $r$  de Pearson.

Descriptores Acústicos		Variables de Entrada
$L_{Aeq}$		Caudal de Motocicletas-Ciclomotores (0.645), Caudal de Vehículos Ligeros (0.514), Caudal de Autobuses (0.433), Caudal de Vehículos Pesados (0.425), Actividades Comerciales/Ocio (0.405)
$L_{eq}$		Caudal de Autobuses (0.645), Caudal de Vehículos Ligeros (0.508), Caudal de Motocicletas-Ciclomotores (0.504), Caudal de Vehículos Pesados (0.478), Tipo de Día (-0.417)
$L_{eq}-L_{Aeq}$		Número de Vehículos con Sirena (-0.283), Velocidad Media (-0.266), Pendiente de Circulación (0.238)
TSLV		Tiempo de Estabilización (0.840), Presencia de Eventos Sonoros Anómalos (0.582), Número de Vehículos con Sirena (0.478), Pendiente de Circulación (0.448)
CF		Presencia de Eventos Sonoros Anómalos (0.827), Número de Vehículos con Sirena (0.623), Tiempo de Estabilización (0.514), Caudal de Motocicletas-Ciclomotores (0.412)
Bandas de 1/3 de octava	20-25 Hz	Caudal de Autobuses (0.264), Caudal de Vehículos Pesados (0.227), Tipo de Pavimento (0.226), Pendiente de Circulación (0.207).
	31,5-125 Hz	Caudal de Autobuses (0.534), Caudal de Vehículos Pesados (0.486),

		Caudal de Vehículos Ligeros (0.447), Caudal de Motocicletas-Ciclomotores (0.431), Tipo de Pavimento (0.366).
	160-2500 Hz	Caudal de Autobuses (0.549), Caudal de Motocicletas-Ciclomotores (0.543), Caudal de Vehículos Ligeros (0.492), Caudal de Vehículos Pesados (0.482), Número de Vehículos con Sirena (0.426), Actividades Comerciales/Ocio (0.405)
	3150-20000 Hz	Caudal de Motocicletas-Ciclomotores (0.499), Actividades Comerciales/Ocio (0.426), Caudal de Vehículos Ligeros (0.408), Caudal de Autobuses (0.356), Caudal de Vehículos Pesados (0.319).

En la tabla 2, podemos observar las variables con un mayor nivel de correlación con respecto a los diferentes indicadores acústicos utilizados en este trabajo. En primer lugar, observamos que para los descriptores  $L_{Aeq}$  y  $L_{eq}$ , el caudal de motocicletas-ciclomotores ( $r = 0.645$ ) y el caudal de autobuses ( $r = 0.645$ ), son las variables de entrada con un mayor nivel de correlación respectivamente. El número de vehículos con sirena ( $r = -0.283$ ) es la variable de entrada con el descriptor  $L_{eq}-L_{Aeq}$ . Por su parte, las variables tiempo de estabilización ( $r = 0.840$ ) y presencia de eventos sonoros anómalos ( $r = 0.827$ ), son las variables con mayor grado de correlación con los descriptores para la caracterización temporal del nivel sonoro, TLSV y CF, respectivamente. Por lo que respecta a las bandas de 1/3 de octava (20 Hz – 20 KHz), se observan diferencias en función del rango de frecuencias en que nos situemos. En el rango 20-25 Hz, 31.5-125 Hz y 160-2500 Hz, el caudal de autobuses es la variable con un mayor nivel de correlación ( $r = 0.264$ ,  $0.534$  y  $0.549$ , respectivamente). Mientras, en el rango 3.15-20 KHz, el caudal de motocicletas-ciclomotores tiene el mayor grado de correlación ( $r = 0.449$ ).

### 3.2 Resultados del modelo de predicción basado en redes neuronales.

En las tablas 3 y 4, podemos observar los resultados obtenidos a partir del modelo de predicción propuesto, basado en redes neuronales, para cada uno de los diferentes descriptores acústicos utilizados. En función de los rangos de valores de los parámetros acústicos (mostrados en la tabla 1), comprobamos que la red neuronal propuesta alcanza una elevada precisión en la predicción de los indicadores sonoros. Los descriptores que son estimados con mayor exactitud (tabla 4) son los parámetros  $L_{Aeq}$ ,  $L_{eq}$  y TSLV, mientras que los descriptores CF, bandas de 1/3 de octava (20 Hz – KHz) y  $L_{eq}-L_{Aeq}$  son estimados con un nivel de precisión ligeramente inferior.

Tabla 3. Error medio de predicción (MPE) y valor  $R^2$  de los conjuntos de entrenamiento.

Descriptores Acústicos	Conjunto 1		Conjunto 2		Conjunto 3		Conjunto 4		Conjunto 5	
	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$
$L_{Aeq}$ [dB(A)]	0.71±0.64	0.96	0.70±0.67	0.96	0.71±0.69	0.96	0.68±0.62	0.97	0.70±0.63	0.97
$L_{eq}$ [dB]	0.86±0.80	0.92	0.94±0.81	0.91	0.89±0.76	0.92	0.88±0.78	0.92	0.89±0.79	0.92
$L_{Aeq}-L_{eq}$	0.85±0.78	0.83	0.90±0.80	0.82	0.86±0.78	0.83	0.78±0.72	0.86	0.83±0.76	0.84
TSLV [dB <sup>2</sup> ]	0.91±0.84	0.98	0.92±0.87	0.98	0.89±0.88	0.98	0.92±0.90	0.98	0.91±0.92	0.98
CF	0.15±0.17	0.86	0.15±0.17	0.86	0.15±0.16	0.87	0.14±0.16	0.87	0.14±0.16	0.87
Bandas de 1/3 de octava [Hz]	1.54±1.46	0.89	1.59±1.46	0.89	1.57±1.48	0.89	1.51±1.44	0.90	1.52±1.45	0.90

Tabla 4. Error medio de predicción (MPE) y valor  $R^2$  de los conjuntos de validación.

Descriptores Acústicos	Conjunto 1		Conjunto 2		Conjunto 3		Conjunto 4		Conjunto 5	
	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$
$L_{Aeq}$ [dB(A)]	0.82±0.76	0.93	0.86±1.05	0.93	0.85±0.95	0.94	0.83±0.92	0.95	0.90±0.81	0.95
$L_{eq}$ [dB]	0.90±0.73	0.91	1.08±0.84	0.90	0.97±0.85	0.91	1.05±0.89	0.91	1.04±0.86	0.91
$L_{Aeq}-L_{eq}$	0.86±0.70	0.82	1.05±0.95	0.78	0.92±0.83	0.82	1.02±0.93	0.82	0.97±0.91	0.80
TSLV [dB <sup>2</sup> ]	0.96±1.05	0.98	1.14±1.25	0.97	1.10±1.15	0.97	1.18±1.12	0.97	1.07±1.18	0.96
CF	0.15±0.15	0.85	0.16±0.18	0.87	0.15±0.16	0.89	0.16±0.16	0.89	0.14±0.15	0.89
Bandas de 1/3 de octava [Hz]	1.70±1.69	0.83	1.79±1.86	0.85	1.76±1.78	0.86	1.72±1.73	0.87	1.76±1.72	0.86

Como observamos en la figura 2, el factor de correlación  $R^2$  varía dependiendo de la banda de 1/3 de octava considerada. Las bandas con un mayor valor del factor  $R^2$  están localizadas entre los 125-3150 Hz, mientras que las bandas con muy menor grado de correlación están situadas entre 20-31.5 Hz, 80-100 Hz and 16-20 KHz. Sin embargo, observando el valor medio del factor  $R^2$  de los 5 conjuntos de datos, verificamos que el valor mínimo es 0.77, alcanzando valores de 0.92.

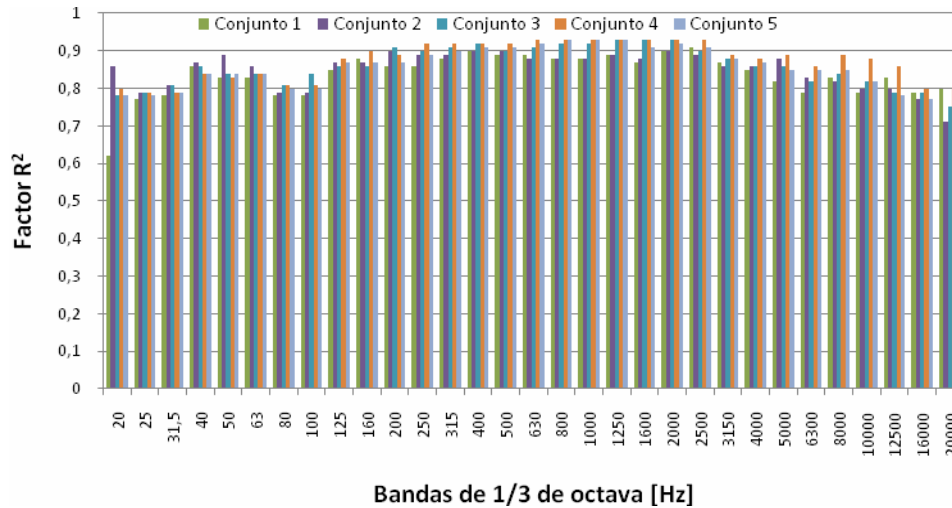


Figura 2. Valor  $R^2$  de las bandas de 1/3 de octava (20-20000 Hz) para los conjuntos de validación.

En vista de los resultados obtenidos, podemos establecer que la red neuronal propuesta es capaz de predecir con una considerable precisión y exactitud el nivel de presión sonora (ponderado A y no ponderado), así como la composición temporal y espectral de los diferentes tipos de situaciones presentadas a la red, situaciones que incluyen una gran heterogeneidad y complejidad (características de las aglomeraciones urbanas).

### 3.3 Comparación con los resultados obtenidos mediante regresión lineal múltiple.

Una vez analizada la precisión y exactitud del modelo de predicción basado en redes neuronales, hemos comparados los resultados obtenidos con los resultados que obtendríamos si utilizáramos una regresión lineal múltiple (a partir de las mismas variables de entrada), en su lugar. Observando los resultados mostrados en la tabla 5, podemos establecer que la utilización de una red neuronal supone una considerable mejora en cuanto a la consecución de un mayor grado de precisión y exactitud en la estimación de los parámetros acústicos utilizados, especialmente para los parámetros  $L_{eq}$ - $L_{Aeq}$  y bandas de 1/3 de octava.

Tabla 5. Comparación de los resultados obtenidos mediante la red neuronal propuesta y una regresión lineal múltiple.

Descriptores Acústicos	Entrenamiento				Validación			
	Red Neuronal		Regresión Lineal Múltiple		Red Neuronal		Regresión Lineal Múltiple	
	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$	MPE	$R^2$
$L_{Aeq}$ [dB(A)]	0.76±0.72	0.96	1.92±1.80	0.67	0.87±0.90	0.94	1.96±1.78	0.71
$L_{eq}$ [dB]	0.92±0.83	0.91	1.70±1.43	0.69	1.01±0.89	0.91	1.79±1.47	0.71
$L_{Aeq}$ - $L_{eq}$	0.85±0.78	0.84	1.50±1.17	0.51	1.03±0.99	0.81	1.96±1.56	0.29
TSLV [dB <sup>2</sup> ]	0.90±0.87	0.98	2.97±2.93	0.77	0.92±1.00	0.97	2.84±2.56	0.75
CF	0.14±0.16	0.87	0.16±0.21	0.75	0.16±0.16	0.88	0.19±0.30	0.75
Bandas de 1/3 de octava [Hz]	1.59±1.51	0.90	2.92±2.67	0.61	1.82±1.78	0.86	3.15±2.83	0.56

Por lo que respecta al valor del factor  $R^2$  de las distintas bandas de 1/3 de octava, obtenido mediante la utilización de la red neuronal propuesta y una regresión lineal múltiple (figura 3), comprobamos que la red neuronal estima con mayor precisión todo el espectro sonoro. Por otro lado, al igual que en el caso de la utilización de la red neuronal propuesta (figura 2), observamos que utilizando una regresión lineal múltiple, las bandas con mayor valor del factor  $R^2$  son las comprendidas entre 125-3150 Hz, sin embargo, en este caso la estimación de las muy bajas y las muy altas frecuencias se realiza con un nivel de precisión muy bajo.

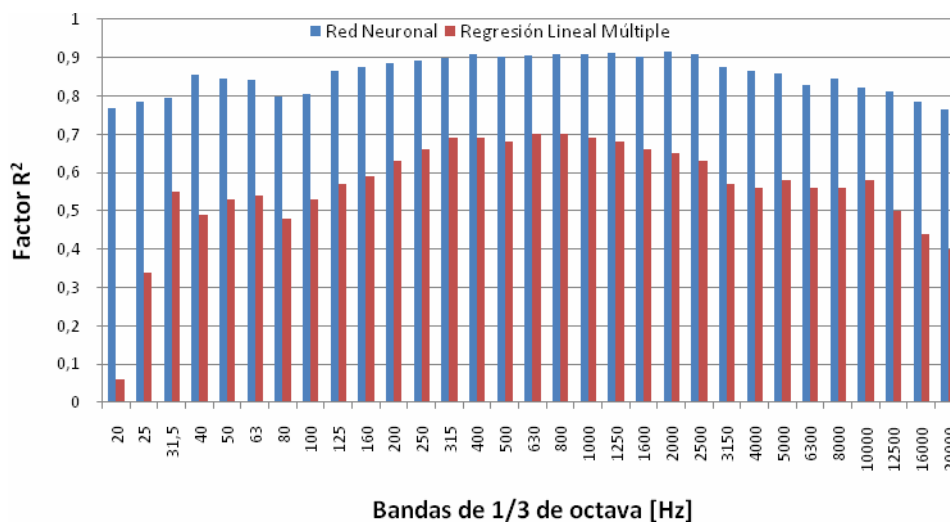


Figura 3. Comparación de los resultados obtenidos mediante la red neuronal propuesta y una regresión lineal múltiple en cuanto al valor  $R^2$  de las bandas de 1/3 de octava (20-20000 Hz) para los conjuntos de validación.

## CONCLUSIONES

En vista de los resultados obtenidos, podemos concluir que la red neuronal propuesta logra una buena predicción de la composición temporal y espectral, así como del nivel de presión sonora (ponderado A y no ponderado) con una gran precisión y exactitud. Esta metodología y los resultados obtenidos, nos permiten ser optimistas con la posibilidad de tener una herramienta para la predicción del nivel de presión sonora y de la composición temporal y espectral, la cual pueda ayudar a la integración de las variables acústicas en la planificación urbana, con el objetivo final de adaptar los paisajes sonoros urbanos a la población que interactúa con ellos.

## REFERENCIAS

- [1] R.M. Schafer: The soundscape: Oursonic environment and the tuning of the world. Destiny Books, Rochester, Vermont, USA, 1994.
- [2] D. Botteldooren, B. De Coensel, T. De Muer: The temporal structure of urban soundscapes. Journal of Sound and Vibration, 292 (2006) 105-123.
- [3] G. Cammarata, S. Cavalieri, A. Fichera: A neural network architecture for noise prediction. Neural Networks, 8 (1995) 963-973.
- [4] A. J. Torija, D.P. Ruiz, D. Botteldooren, B. De Coensel: Taking into account the temporary and spectral structure of the sound energy for the characterization of the annoyance generated by the road and railway traffic. 5<sup>th</sup> Congreso Ibérico de Acústica (Acústica '08), Coimbra, Portugal, 2008.