

# PRIMERAS JORNADAS REGIONALES DE ACÚSTICA AdAA 2009

19 y 20 de noviembre de 2009, Rosario, Argentina



AdAA2009-A056R

## Representación de señales de audio con descomposición empírica de modos y submuestreo adaptativo

Fernando A. Marengo Rodríguez<sup>(a)</sup>,  
Federico Miyara<sup>(b)</sup>.

(a), (b) Laboratorio de Acústica y Electroacústica, Escuela de Ingeniería Electrónica, Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Riobamba 245 bis, Rosario, Argentina. E-mail: (a) [fmarengorodriguez@yahoo.com.ar](mailto:fmarengorodriguez@yahoo.com.ar), (b) [fmiyara@fceia.unr.edu.ar](mailto:fmiyara@fceia.unr.edu.ar)

### Abstract

Lossless audio coding importance has been increasing in the last years. To this aim, many algorithms have been developed and some of them have become international standards. Despite their effectiveness, these methods assume linearity on the signal under study, which is not always true since acoustic signals are highly affected by nonlinearities. In this paper, a novel representation of acoustic signals is presented by using the empirical mode decomposition method, which has been proved suitable for nonlinear signals analysis. Subsequent subsampling of the sequences obtained from this decomposition is performed, giving a reduced data set. This method is applied to a short time signal and its performance is evaluated by means of the resulting compression factor.

### Resumen

En los últimos años se ha incrementado la importancia de la compresión de audio sin pérdidas. Para tal fin se han desarrollado numerosos algoritmos, y algunos de ellos se han convertido en estándares internacionales. A pesar de la buena eficiencia obtenida, estas metodologías suponen linealidades en las señales analizadas, lo cual no siempre es cierto si se tiene presente que las mismas son altamente no lineales. En este trabajo se presenta una nueva representación de señales acústicas por medio de la descomposición empírica de modos, la cual se corroboró que es adecuada para el análisis de señales no lineales. Posteriormente, se submuestra cada señal obtenida en la descomposición a tasa suficientemente baja, resultando un pequeño conjunto de datos. Este método se aplica a una señal de corta duración temporal, y su performance se evalúa mediante el factor de compresión resultante.

## 1 Introducción

La compresión de información es de vital importancia tanto para el almacenamiento como para el traslado de datos. La teoría de la información introducida por Shannon demuestra que todo conjunto de símbolos en un lenguaje dado posee redundancia que al ser suprimida permite comprimir los datos de entrada. Este hecho es conocido como codificación de fuente. Otro concepto también relevante para el traslado de la información es la codificación del canal, que se vincula con el agregado de redundancia controlada a los datos de interés a fin de que éstos se puedan enviar exitosamente por canales de comunicación imperfectos. El presente trabajo está abocado a un tipo particular de codificación de fuente: el audio. Se supone que los datos a codificar provienen en formato wav con cuantificación uniforme de  $n_b$  bits.

Hasta el momento se han desarrollado numerosos métodos de compresión de audio. Por un lado, los métodos de compresión sin pérdidas conservan la forma de onda de entrada y se basan en la estimación del espectro de la señal de entrada mediante filtros apropiados. Por otra parte se desarrollaron algoritmos de compresión perceptual tales como Ogg Vorbis y MP3. Éstos consisten en la remoción de redundancias así como de irrelevancias psicoacústicas, logrando una mayor reducción de la información a cambio de distorsiones no audibles o poco apreciables en la forma de onda.

Centramos nuestro estudio en la compresión sin pérdidas, aunque bien podría extenderse este estudio a la compresión perceptual vía la aplicación de modelos psicoacústicos apropiados.

Este trabajo se organiza de la siguiente manera. Primero se explica brevemente el funcionamiento de los métodos de compresión de audio sin pérdidas, luego se expone el algoritmo de descomposición empírica de modos y finalmente se presenta el método propuesto para la representación económica de una señal de audio dada.

## 2 Métodos de compresión existentes

Las técnicas de compresión sin pérdidas más conocidas actualmente como por ejemplo FLAC (Xiph.Org, 2009) y Monkey's Audio (Ashland, 2009) se basan en la predicción de los valores de la secuencia de entrada a través de un estimador. Dicha estimación se realiza en cada instante como una combinación lineal de valores previos de la misma señal de entrada. Como resultado, el estimador brinda un valor numérico que se aproxima al presente en la entrada. El error de esta estimación, llamado residuo, posee la propiedad de que es más fácilmente compactable (o codificable) que la secuencia de entrada. Como resultado, la compresión se basa en la extracción de los coeficientes empleados en el estimador y en la codificación del residuo. Para aplicaciones típicas se economiza el 50 % de los bits de entrada. Dicho de otra manera, el factor de compresión (la relación entre los tamaños del archivo original y del comprimido) es 2.

Es claro que estos métodos de codificación suponen linealidades en la entrada, dada la naturaleza del predictor empleado. Sin embargo, se sabe que las señales acústicas (como por ejemplo las que representan una nota musical) poseen múltiples componentes con modulaciones de amplitud (AM) y frecuencia (FM), lo que nos lleva a buscar una técnica de representación de señales más acorde a dicha naturaleza.

## 3 Algoritmo de descomposición empírica de modos

El algoritmo de descomposición empírica de modos, más conocido como EMD (Huang et al., 1998) es un método desarrollado para el análisis de señales reales no estacionarias y no

lineales. Esta técnica se basa en la extracción sucesiva de las oscilaciones presentes en el estímulo de entrada, desde el detalle más fino hasta el detalle más grueso. La aplicación de esta metodología a una señal dada brinda un reducido conjunto de funciones de tipo AM-FM con media nula. Éstas son llamadas funciones de modo intrínseco o IMF.

El funcionamiento del algoritmo EMD es el siguiente:

- 1 - Se extraen los máximos locales y mínimos locales de la señal de entrada  $x(t)$ .
- 2 - Se calcula la envolvente superior  $e_{\text{sup}}(t)$  interpolando los máximos locales. Se procede en forma análoga con los mínimos locales, obteniéndose la envolvente inferior  $e_{\text{inf}}(t)$ .
- 3 - Se calcula el valor medio local como el promedio de las envolventes calculadas en el paso anterior, es decir  $m(t) = [e_{\text{inf}}(t) + e_{\text{sup}}(t)] / 2$ .
- 4 - Se suprime la media local a la señal de entrada, obteniéndose la función detalle  $d(t) = x(t) - m(t)$  que a priori es considerada IMF. Por otro lado se calcula el residuo como  $r(t) = x(t) - d(t)$ .
- 5 - Se repiten los pasos 1 a 4 tomándose como estímulo de entrada el residuo en lugar de  $x(t)$ .

En realidad, la función detalle extraída en el paso 4 no es simétrica alrededor del cero, por lo cual su valor medio es no nulo. Para minimizar ese error, se itera entre los pasos 1 y 4 usando como entrada  $d(t)$  en lugar de  $x(t)$  hasta que su valor medio es menor que un umbral predeterminado. Cuando se cumple esta condición, el detalle  $d(t)$  es considerado IMF<sup>1</sup>.

Cabe aclarar que los pasos 1 a 5 se llevan a cabo hasta que el residuo es o bien muy pequeño o bien una función monótona de la cual no pueden rescatarse más IMF (es decir, no se pueden extraer más oscilaciones del estímulo). Como resultado, el análisis por EMD da lugar a la siguiente identidad:

$$x(t) = \sum_{k=1}^K IMF_k(t) + r(t), \quad (1)$$

donde  $K$  es la cantidad total de IMF y  $r(t)$  es una función monótona llamada residuo final o resto. Claramente, la secuencia de entrada no se expresa en términos de una combinación lineal, sino más bien como una suma de componentes no lineales en vista de la presencia de modulación en frecuencia presente en cada una de ellas.

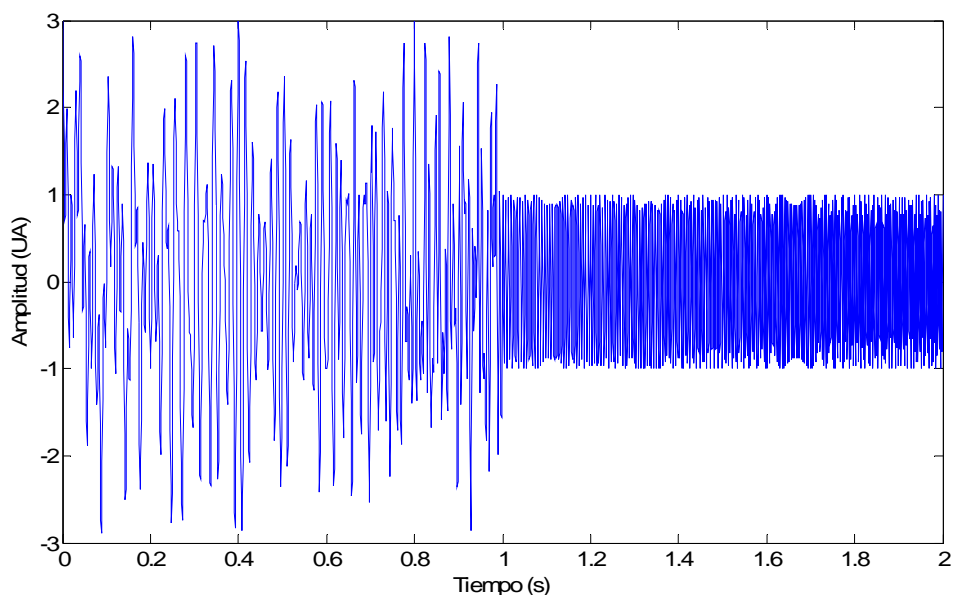
### 3.1 Ejemplo de análisis

A modo de ejemplo, se analiza con EMD un chirp compuesto definido en el intervalo temporal  $[0,2]$  s, muestreado a una tasa  $f_s = 1$  kHz y cuantificado con  $n_b = 8$  bits. Dicha señal se muestra en la figura 1 y consiste en la suma de tres chirp cuadráticos: uno de 2 s de duración y frecuencia creciente entre cero y 300 Hz, los otros chirp se desarrollan en el primer segundo y evolucionan de 25 Hz a 100 Hz, y de 100 Hz a 25 Hz respectivamente. Cabe aclarar que se propone esta señal a fin de contrastar los resultados aquí obtenidos con los expuestos en el trabajo de A. Linderhed (Linderhed, 2002).

El estímulo se expresa en formato de punto flotante y se analiza con EMD, resultando las IMF mostradas en la figura 2. Como se observa, resulta un pequeño conjunto de funciones AM-FM con media nula y tasa de oscilación decreciente a medida que aumenta el orden de la IMF. Otro aspecto importante a considerar es que la cantidad de funciones IMF detectadas depende de la cantidad de muestras  $N_t$  del estímulo, así como de su riqueza armónica. Cuando

<sup>1</sup> En este trabajo se utilizó el algoritmo de EMD disponible en <http://perso.ens-lyon.fr/patrick.flandrion/emd.html>

se analiza ruido blanco con EMD, la cantidad de IMF es aproximadamente  $K = \log_2 N_t$  (Wu y Huang, 2004). Este límite se respeta en el ejemplo numérico aquí estudiado.



**Figura 1.** Señal analizada.

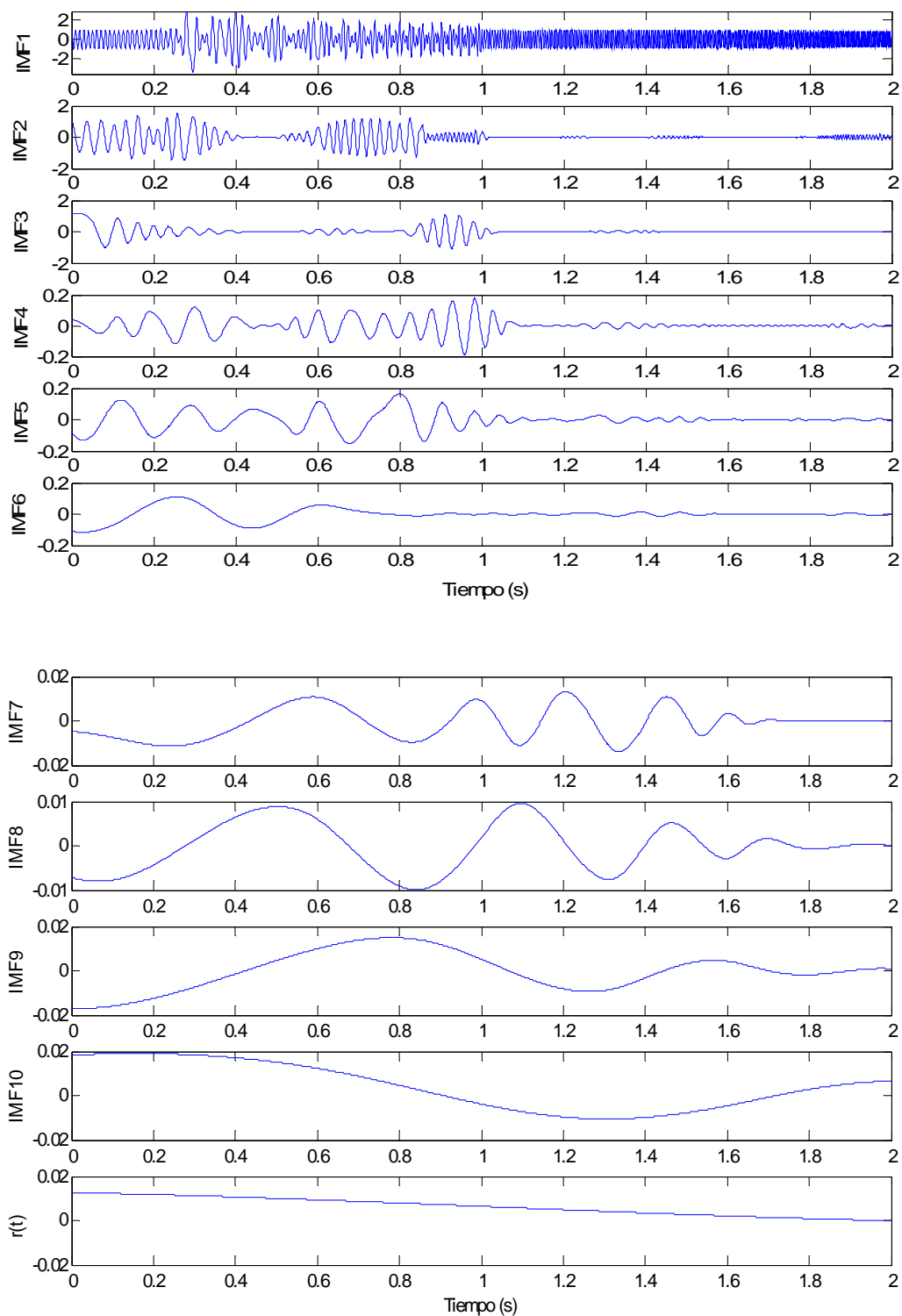
#### **4 Método propuesto**

El método propuesto en este trabajo consiste en la implementación del diagrama de la figura 3. Comienza con el análisis de la señal de entrada vía EMD. Posteriormente, las señales moduladas son submuestreadas a la mínima tasa posible sin pérdida de información, y las muestras rescatadas son codificadas y multiplexadas para su envío.

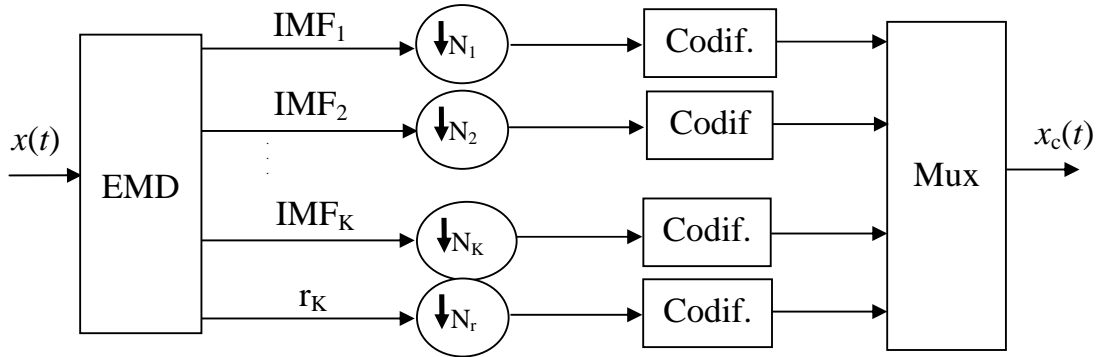
Aunque no se muestra explícitamente en la figura 3, la señal de estímulo es recuantificada a una alta resolución para su eficiente análisis por EMD. Además, en el presente caso se usa un codificador de entropía sin memoria, y los datos enviados a la salida incluyen los valores de frecuencias usados en los moduladores de fase. A continuación se explicarán los detalles pertinentes a los bloques del esquema propuesto.

#### **5 Agrupamiento de funciones IMF**

Debido a que EMD se aplica a señales de alta resolución, las IMF extraídas pueden poseer amplitudes comparables o menores que el ruido de cuantificación de la señal de entrada. En el ejemplo de la sección 3.1, la resolución de entrada (es decir la mínima amplitud discriminable) es de 0,0118 y comparable a la amplitud de las IMF de mayor orden (ver figura 2). Si se desea cuantificar cada una de estas funciones con la misma resolución de entrada, se acumularán errores y se perderá información.



**Figura 2.** Componentes de la señal de entrada detectadas con EMD.



**Figura 3.** Esquema propuesto para la representación económica de señales de audio.

En esta sección se propone agrupar en una sola componente las IMF de amplitud comparable a la resolución adoptada con  $n_b$  bits. Para este procedimiento, se considera que la función  $IMF_k(t)$  es correctamente representable por su versión cuantificada  $IMF_{kq}(t)$  siempre que el error de cuantificación esté suficientemente acotado, es decir

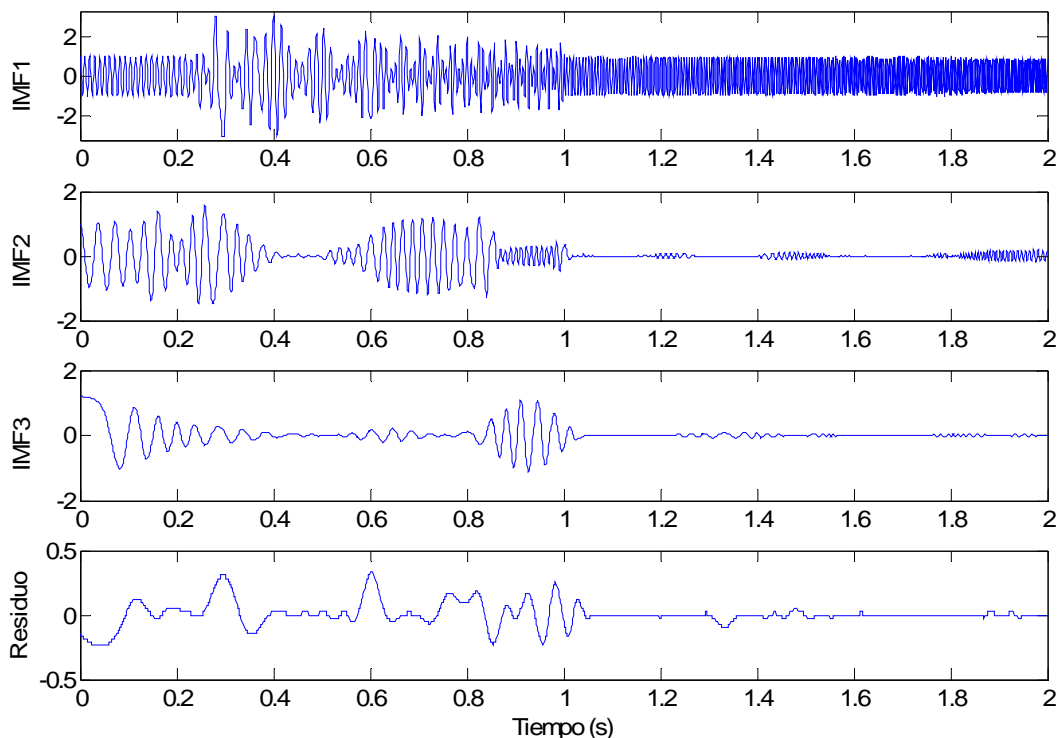
$$\frac{\text{rms}\{IMF_k(t) - IMF_{kq}(t)\}}{\text{rms}\{IMF_k(t)\}} < \mu, \quad (2)$$

donde  $\mu$  es un umbral predeterminado elegido empíricamente y  $\text{rms}\{y(t)\}$  es el valor eficaz de la secuencia  $y(t)$ . Al adoptar  $\mu = 0,1$  para el chirp compuesto resultó un agrupamiento de las IMF de orden 4 en adelante incluyendo el resto. En la figura 4 se muestra el resultado obtenido con las IMF cuantificadas a fin de ilustrar la eficiencia del agrupamiento.

## 6 Submuestreo y reconstrucción

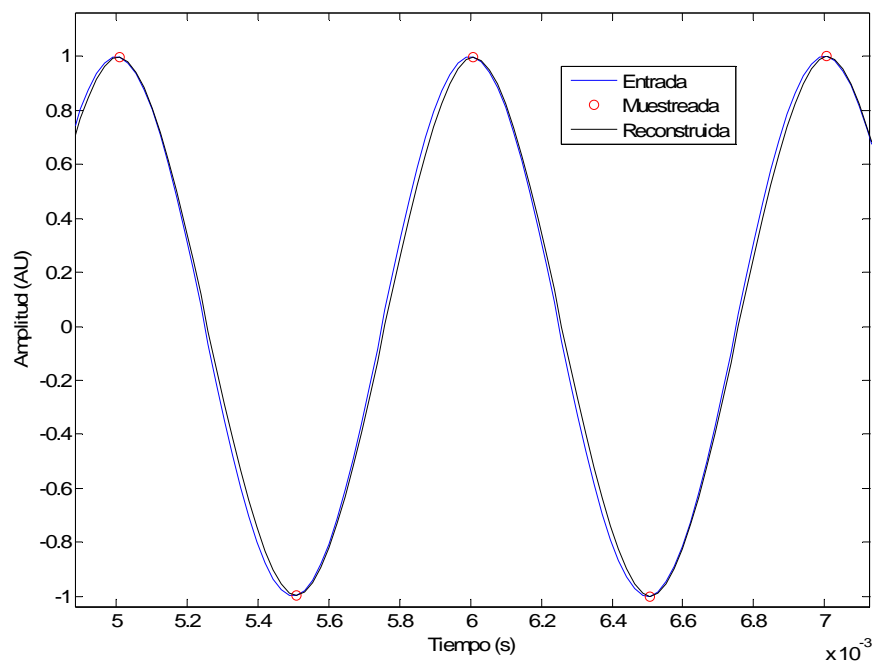
En estudios anteriores se demostró que EMD funciona como un banco de filtros diádicos (Flandrin et al., 2004), es decir que equivale a un filtro de octavas aplicado *localmente* y en forma recursiva. En el caso que la señal estudiada sea ruido blanco, las IMF poseen el mismo ancho de banda en una escala semilogarítmica (Wu y Huang, 2004). Dado que las IMF dependen solamente de la riqueza armónica de la entrada, sus anchos de banda no son conocidos *a priori*, pero sí *a posteriori*, es decir luego del análisis con EMD. Entonces, cada una de estas funciones es esencialmente de tipo pasabanda (PBn) con ancho espectral dependiente del estímulo de entrada. En particular, la componente que resulta del agrupamiento de IMF (nuevo residuo) también es PBn pero con mayor ancho de banda que las IMF de las cuales proviene.

Por otro lado, el teorema del muestreo garantiza que una señal analógica de ancho de banda  $B$  Hz puede ser representada sin ambigüedad a partir de sus muestras discretas equiespaciadas  $T_s = 1/f_s$  s siempre que se cumpla  $f_s > B/2$  (Marks, 1991).



**Figura 4.** IMF agrupadas y cuantificadas con  $n_b$  bits para economizar el análisis del estímulo. Nótese que el nuevo residuo posee valores comparables a la resolución de entrada.

En los métodos de compresión de audio existentes se suele muestrear a tasa crítica (Spanias et al., 2007), es decir que se adopta el caso límite  $f_s = B/2$ . Si se adopta este criterio en el presente trabajo, se recuperará la señal de entrada con pequeña distorsión ya que por un lado no se respeta la desigualdad  $f_s < B/2$ , y además no se interpola con una función sinc, sino con otra de soporte compacto. Un ejemplo de tal aplicación usando interpolación spline (que es la empleada en todo este documento) se ilustra en la figura 5. Como allí se observa, se muestrearon los máximos y mínimos locales presentes en la entrada, y la señal reconstruida está levemente desfasada de la original.



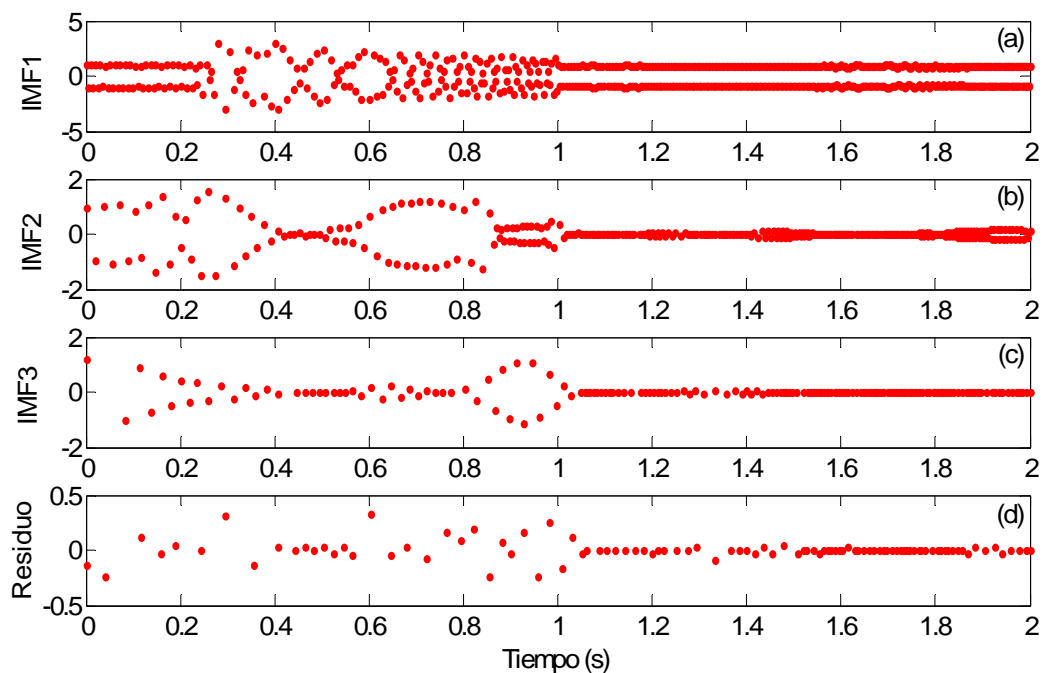
**Figura 5.** Armónica de entrada, sus muestras extraídas a tasa crítica y armónica reconstruida.

### 6.1 Aplicación al método presente

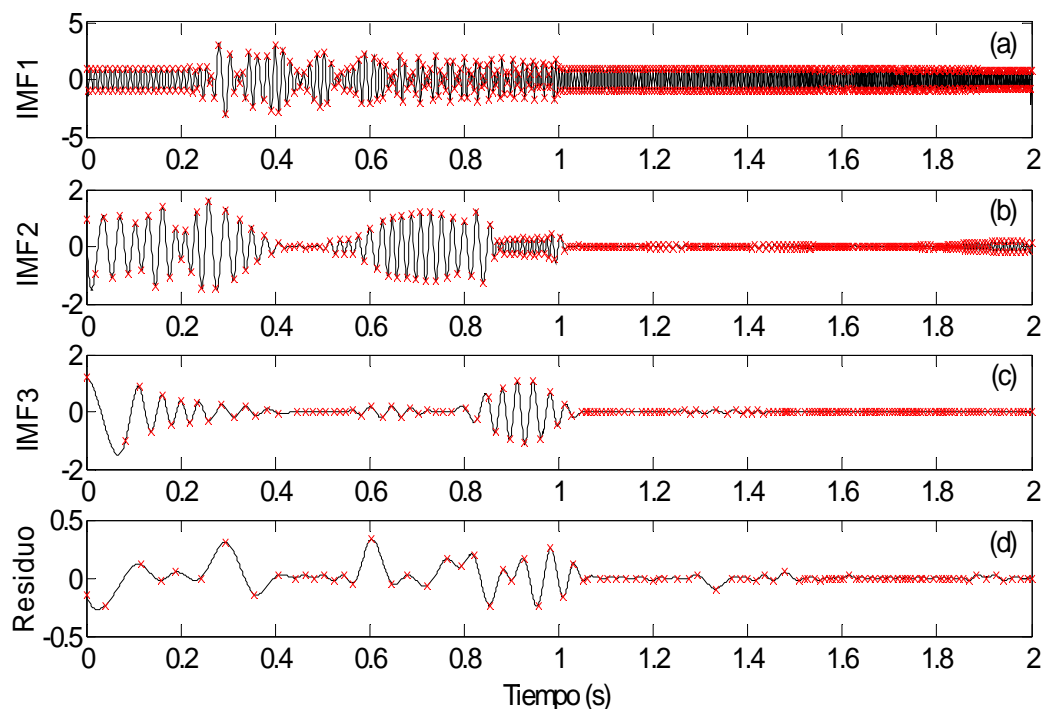
Se propone submuestrear y cuantificar con la misma resolución de entrada los extremos locales de cada función de modo intrínseco, obteniéndose un conjunto de puntos  $P_1$ . Suponiendo que dicha función es analógica, el submuestreo equivale al muestreo a tasa crítica *local*, ya que a priori podría recuperarse la IMF a partir de sus extremos locales. Este proceso fue aplicado a imágenes (Linderhed, 2002), pero aquí se lo aplica al audio y con mejoras.

Al aplicarse el submuestreo propuesto a las funciones presentadas en la figura 4 se logran los resultados expuestos en la figura 6. Del total de 2001 muestras de la entrada, se rescatan 1244. Para corroborar la precisión de este submuestreo, se reconstruyó cada IMF vía interpolación de los puntos  $P_1$ . Los resultados gráficos de dicha reconstrucción y del error absoluto producidos por esta causa se presentan en la figura 7 y 8 respectivamente. Como se observa en esas gráficas, el mencionado error se puede atribuir a que no se respeta *localmente* el límite  $f_{sl} < B_1/2$  entre puntos  $P_1$  consecutivos.  $B_1$  es el ancho de banda existente entre dichos puntos y  $f_{sl}$  es la tasa de muestreo asociada al intervalo temporal entre ellos.

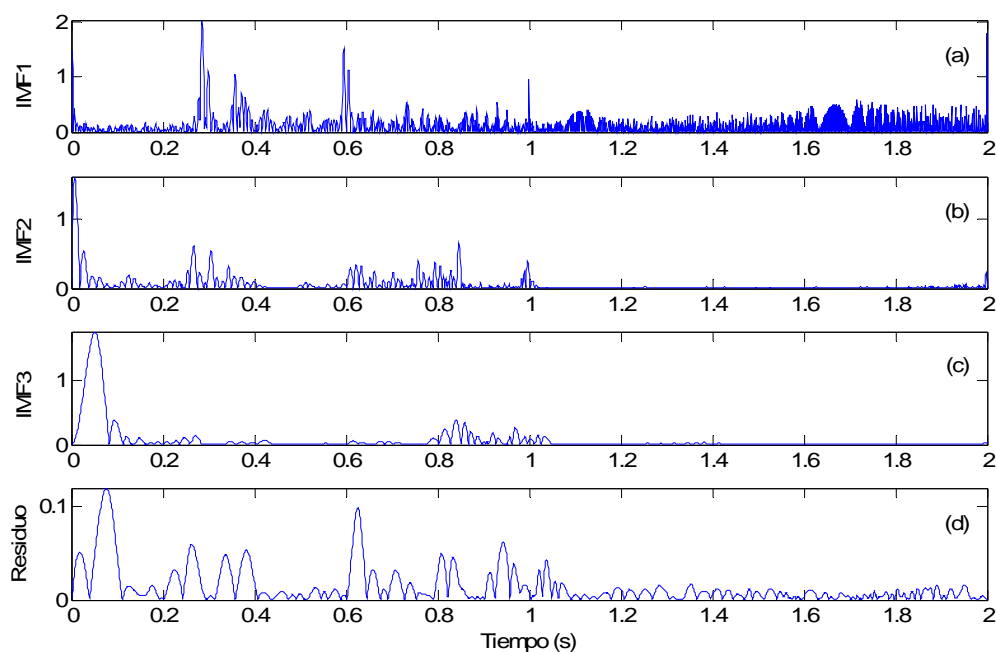




**Figura 6.** Componentes IMF y residuo submuestreados y cuantificados en sus extremos locales.

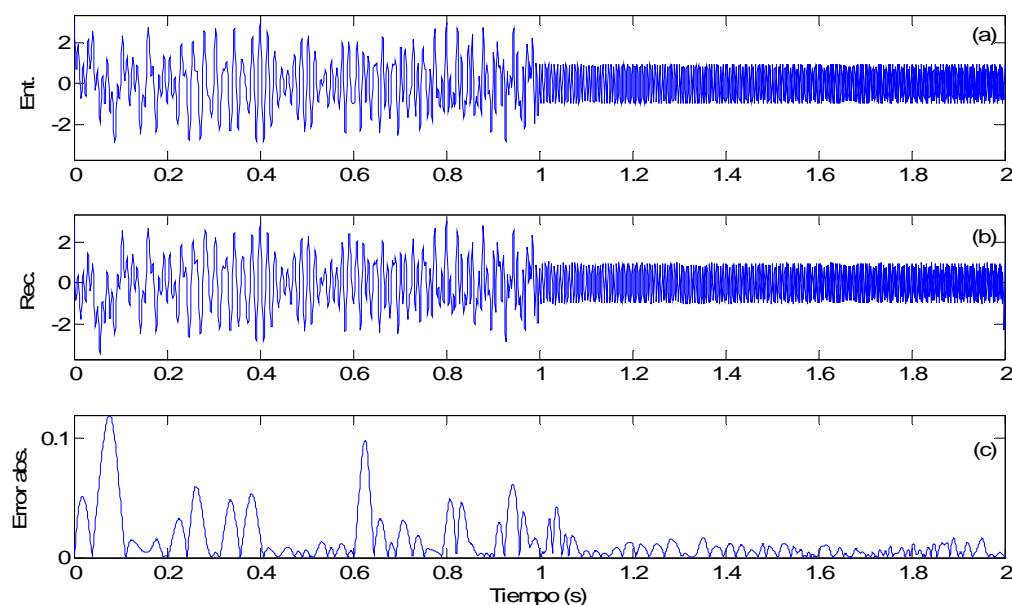


**Figura 7.** Funciones IMF reconstruidas (en trazo negro continuo) vía interpolación de los extremos extraídos de las funciones originales (en cruces rojas). Nótese el error de reconstrucción traducido en desfases entre los máximos originales y los de las funciones reconstruidas.



**Figura 8.** Error absoluto producido por la reconstrucción de cada función de modo intrínseco y del residuo.

Finalmente, la señal de entrada se recupera a través de la suma de las IMF reconstruidas. En la figura 9 panel (a) se muestra el chirp de entrada, en el panel (b) el reconstruido y en el panel (c) el error absoluto total. El estímulo de entrada se logra recuperar aunque con cierto error. Actualmente los autores de este trabajo están investigando formas de optimizar la reconstrucción en función del submuestreo adoptado.



**Figura 9.** Señal de entrada (a), reconstruida (b) y error absoluto total (c).

## 7 Codificación

Para tener una idea del grado de compactación logrado con el método propuesto, se determinó la entropía de los símbolos extraídos (extremos de las IMF) y la cantidad de bits necesarios para su representación. Por definición, la entropía brinda una medida de incerteza de los símbolos provenientes de una dada fuente de información. Más específicamente, este parámetro permite estimar la mínima cantidad de bits que puede representar cada símbolo en promedio en un codificador sin memoria. Si una secuencia de símbolos provenientes de la fuente consiste en valores numéricos discretos  $a_i$  (dentro del rango uniforme discretizado  $[-2^{n_b-1}, 2^{n_b-1} - 1]$ ) con probabilidad  $p(a_i)$ , se define la entropía  $H$  como (Shannon, 1948)

$$H = - \sum_{a_i} p(a_i) \log_2 p(a_i). \quad (3)$$

Por otro lado, para tener una medida de cuán concentrados están los valores  $a_i$  de la secuencia analizada alrededor de una región determinada, se calculó la entropía relativa  $H_r$  que se define como

$$H_r = H / H_{\max}, \quad (4)$$

donde  $H_{\max}$  es la entropía que se obtendría si los valores  $a_i$  fueran equiprobables, es decir  $H_{\max} = \log_2 N_t$ , siendo  $N_t$  la cantidad de datos analizados.

Finalmente, para los conjuntos de muestras de la entrada y de los extremos se determinó la cantidad de bits a transmitir. La misma se calculó como el producto entre la entropía del conjunto y la cantidad de muestras presentes en ella. Este resultado permitió estimar la reducción de bits lograda y por ende el factor de compresión alcanzado.

### 7.1 Primeros resultados

La tabla 1 resume los valores numéricos de las entropías y de la cantidad de muestras y de bits asociados al chirp compuesto. Cabe aclarar que se aproximaron las funciones de probabilidad de las muestras por sus correspondientes histogramas. Como en cualquier señal de audio típica, el estímulo en el presente caso no es notablemente compactable ya que su entropía es ligeramente menor que la cantidad de bits  $n_b$  que lo representa. Teniendo en cuenta que el estímulo se representa con 2001 muestras  $\times$  8 bits/muestra = 16008 bits, la compresión de este conjunto de datos sin procesamiento ahorra a lo sumo el 11 % de bits, es decir que el factor de compresión  $F_c$  es de 1,12.

Por otro lado, la cantidad de muestras extraídas con la técnica presente es menor a la de entrada, y más de la mitad de su aporte proviene de la IMF de primer orden. Esto es esperable, ya que esta función resume las variaciones más rápidas de la señal independientemente de la naturaleza de la misma. Otra observación interesante surge al observar que las muestras extraídas poseen menor incertidumbre que las de entrada, ya que tienen menor entropía. Esto explica por qué se reduce sensiblemente la cantidad de bits necesarios para representar las IMF y por consiguiente el factor de compresión. El consecuente ahorro de bits es del 63,6 %.

Los resultados de la entropía relativa dicen que los extremos de cada IMF se hallan más concentrados dentro de ciertas regiones que las muestras de entrada, y que dicha concentración aumenta con el orden de la IMF. Esto es esperable ya que en un intervalo temporal determinado es más probable que las variaciones más rápidas de entrada (IMF de

mayor orden) posean amplitudes más diversas que las variaciones más lentas de la misma (IMF de menor orden).

**Tabla 1.** Resultados asociados a la compresión del chirp compuesto.

	$x(t)$	Extremos locales				
		IMF1	IMF2	IMF3	Residuo	Total
$H$ (bits/muestra)	7,13	5,37	4,71	3,36	2,46	-
$H_r$	0,65	0,58	0,56	0,45	0,37	-
Cant. de muestras	2001	639	335	170	100	1244
Cant. de bits	14267	3431	1578	571	246	5826
$F_c$	1,12	-	-	-	-	2,75

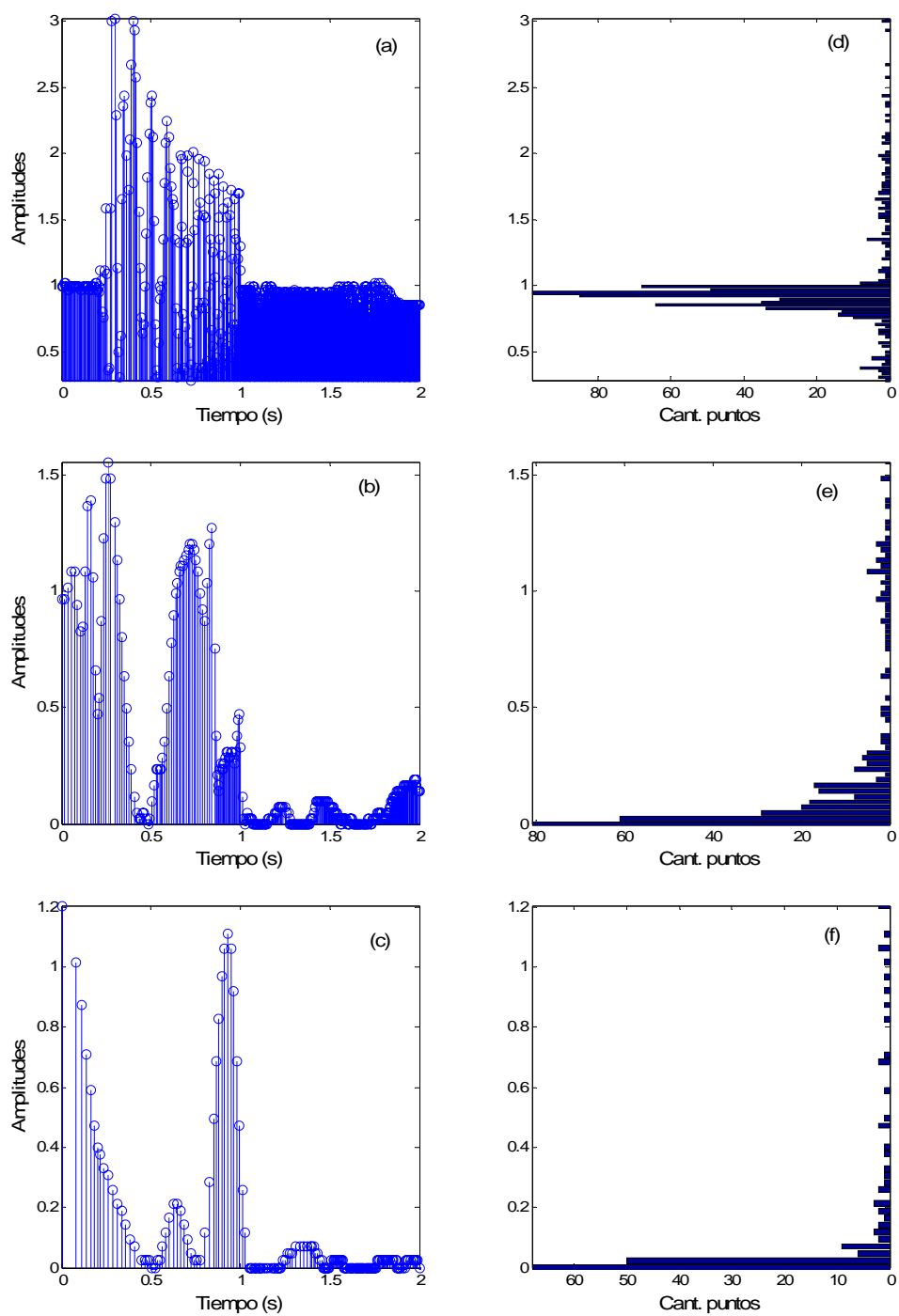
## 7.2 Resultados tomando el valor absoluto de las muestras

Considerando que cada IMF es simétrica alrededor del cero, sus extremos locales son de signo alternante. Se puede entonces ahorrar un bit de información por muestra mediante la representación de su valor absoluto. En el extremo receptor, la información de entrada se recuperaría con un bit adicional en la secuencia correspondiente a cada IMF informando sobre el signo de la primera muestra, de tal forma de invertir alternadamente el signo en las muestras siguientes recibidas. Al adoptar este criterio se obtienen los histogramas y señales temporales graficados en la figura 10. Las entropías, cantidad de bits y factor de compresión obtenidos con esta técnica se exponen en la tabla 2.

**Tabla 2.** Resultados de la evaluación del método tomando el valor absoluto de los extremos de las IMF.

	Extremos locales			
	IMF1	IMF2	IMF3	Total
$H$ (bits/muestra)	4,49	4,06	2,91	-
$H_r$	0,48	0,48	0,39	-
Cant. de bits	2869	1360	495	4970
$F_c$	-	-	-	3,22

Los resultados de la tabla 2 dicen que con respecto al caso anterior, este método optimiza la cantidad de bits a transmitir (y por ende el factor de compresión) en un 15 %. Además, se reduce la entropía relativa debido a una menor dispersión de datos. Es curioso notar en los histogramas de la figura 10 la ausencia de valores de amplitud, lo cual motiva la búsqueda de funciones de transformación necesarias que supriman dichos baches.



**Figura 10.** Valores absolutos de los extremos locales recuperados de la (a) primera IMF, (b) segunda IMF y (c) tercera IMF. Los paneles (d), (e) y (f) muestran los histogramas asociados a las funciones en (a), (b) y (c) respectivamente.

## 8 Conclusiones y trabajo futuro

En este trabajo se propuso una técnica para representar señales digitales de audio de forma económica, y se logró reducir la tasa de bits. Para este fin, se empleó el algoritmo de EMD para descomponer el estímulo en un reducido conjunto de funciones de tipo AM-FM de banda limitada. Luego se aplicó una técnica de agrupamiento de funciones IMF para aquellas de amplitud comparable a la resolución de entrada, lográndose reducir la cantidad de funciones a caracterizar. Posteriormente, se aplicó un método de submuestreo adaptativo que consiste en la extracción de extremos locales, y se adoptaron estos puntos como representativos para cada IMF. Se corroboró este hecho mediante su interpolación, alcanzándose a reconstruir la entrada con poco error. La eficiencia de compresión del método presente se determinó mediante la evaluación del factor de compresión y la cantidad de bits necesarios a transmitir. Esta eficiencia es aún mayor si se elige transmitir las amplitudes sin signo. Los resultados obtenidos acá son preliminares y solo sirven para tener una noción del grado de compresión para un caso particular, sin tener en cuenta el agregado de redundancia propio de todo método de codificación.

Con respecto a la técnica propuesta por A. Linderhed, la presente calcula las entropías sin distorsionar la resolución de entrada y computa la reducción de la cantidad de bits en lugar de la suma de entropías, lo cual no tiene significado relevante para conocer el rendimiento de compresión alcanzado. Otra diferencia con ese trabajo es la propuesta del cómputo de los valores absolutos de los extremos.

Actualmente se está trabajando para introducir mejoras a la técnica aquí expuesta, como por ejemplo las siguientes:

- Reducción del ancho de banda de las IMF mediante optimización del filtrado, lo cual posibilitaría modular cada IMF a banda base y consecuentemente reducir la cantidad de extremos locales obtenidos.
- Diseño de funciones de transformación a los histogramas para suprimir valores de amplitud no existentes en la secuencia de puntos analizada.
- Minimización del error producido en la reconstrucción del estímulo en el receptor.
- Extensión de la metodología propuesta a señales de longitud arbitraria vía el análisis por EMD en cuadros de longitud acotada.

### Referencias

- The Xiph.Org Foundation (2009). "FLAC open-source audio compression program", Internet: <http://flac.sourceforge.net/>
- Ashland, Matthew T. (2009). "Monkey's Audio Compression Program", Internet: <http://www.monkeysaudio.com/>
- Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. C.; Shih, H. H.; Zheng, Q.; Yen, N.-C.; Tung, C. C.; and Liu, H. H. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". *Proceedings of the Royal Society of London, Ser. A*, 454, 903–995.
- Linderhed, Anna (2002). "2D empirical mode decompositions in the spirit of image compression". *Wavelet and Independent Component Analysis Applications IX, Orlando, FL, SPIE Proceedings*, 4738, 1-8.
- Wu, Z. and Huang, N. E. (2004) "A study of the characteristics of white noise using the empirical mode decomposition method". *Proceedings of the Royal Society of London, Ser. A* 460, 1597–1611.
- Flandrin, P.; Rilling, G.; Gonçalves, P. (2004). "Empirical Mode Decomposition as a Filter Bank". *IEEE Signal Processing Letters*, 11, 2, 112-114.
- Marks, R. J. II (1991). "Introduction to Shannon Sampling and Interpolation Theory". Springer-Verlag, New York, USA.

- Spanias, A.; Painter, T.; Atti, V. (2007). "Audio signal processing and coding". John Wiley & Sons, Hoboken, New Jersey, USA.
- Boashash, B. (1992). "Estimating and interpreting the instantaneous frequency of a signal". Proceedings of the IEEE, 80, 4, 520–568.
- Hahn, S. (1996). "Hilbert Transforms in Signal Processing". Artech House, London, UK.
- Bedrosian, E. (1963). "A product theorem for Hilbert transforms," Proceedings of the IEEE, 51, 868–869.
- Shannon, C. E. (1948). "A Mathematical Theory of Communication". The Bell System Technical Journal, 27, 379–423, 623–656.
- Khaldi, K.; Boudraa, A. O.; Turki, M.; Chonavel, T.; and Samaali, I. (2009). "Audio Encoding Based On The Empirical Mode Decomposition". Proceedings of 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland, 924-928.