Codificadores de audio

Fernando A. Marengo Rodriguez, PhD

fmarengorodriguez@yahoo.com.ar

Codificación de Audio Digital

Laboratorio de Acústica y Electroacústica

Escuela de Ingeniería Electrónica

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Universidad Nacional de Rosario, Riobamba 245 bis

www.codecdigital.com.ar

Resumen

- Principios de codificación.
- Codificación de audio.
- Codificación de audio perceptual (con pérdidas).
- Codificación de audio sin pérdidas.
- Tendencias futuras: nuevos codecs.

Principios de codificación

Entropía

- La codificación consiste en reducir la cantidad de bits que emite una fuente de información.
- Si la fuente emite valores numéricos discretos a_i, con probabilidad $p(a_i)$ en el rango $[-2^{n-1}; 2^{n-1}-1]$, se define la **entropía** H como la mínima cantidad de bits con la que se puede representar cada símbolo:

$$H = -\sum_{i=1}^{N_t} p(a_i) \log_2 p(a_i)$$

$H_{máx} = \log_2 N_t$ (distribución uniforme)

Shannon, C. E. (1948). "A Mathematical Theory of Communication". The Bell System Technical Journal, 27, 379–423, 623–656.

Codificación de símbolos

- En **nuestro lenguaje** asignamos:
  - pocas letras a las palabras más frecuentes (“sí”, “ya”, “que”),
  - más letras a las palabras menos frecuentes (“impactante”, “espectacular”, “georeferenciado”).
- **Codificadores estadísticos**: análogamente al lenguaje, se asignan códigos cuya longitud es inversamente proporcional a la frecuencia de aparición de los símbolos emitidos por la fuente.
- **Método basados en diccionarios** (LZ77, LZ78): a medida que se recorren los símbolos del archivo, las cadenas nuevas se almacenan en un diccionario y las cadenas viejas (o sea las repetidas) se asocian al diccionario.

Codificador RAR

- Se basa en el método LZ77 (Lempel Ziv).
- Funcionamiento: se hace una lectura dinámica de los datos con ventanas temporales que abarcan hasta varios kb. Esas ventanas se dividen en 2:
  - una porción larga que abarca los datos ya codificados (ej. 4096 kbytes),
  - una porción corta donde están los nuevos datos a comprimir. Éstos se asocian con porciones de datos ya comprimidos (de la porción larga), que operan a modo de diccionario.

Codificadores de audio

- Los métodos estadísticos no son eficientes, excepto en pocos casos. Las razones:
  - No se explotan las redundancias entre muestras consecutivas, ya que en las señales de audio hay una fuerte dependencia entre ellas (mayor concentración energética en bajas frecuencias)
    - → **redundancia intracanal**
  - No se aprovecha la correlación entre los canales de un archivo de audio
    - → **redundancia intercanal**

Codificadores de audio

Perceptuales

Sin pérdidas (CSP)

Minimizan redundancia en la señal de entrada

- Filtran irrelevancias psicoacústicas → suprimen información
- Gran compresión ~ 8 a 13 veces
- Ejemplos: MPEG-1 audio layer 3 (MP3) y **Ogg Vorbis** (formato libre y abierto).

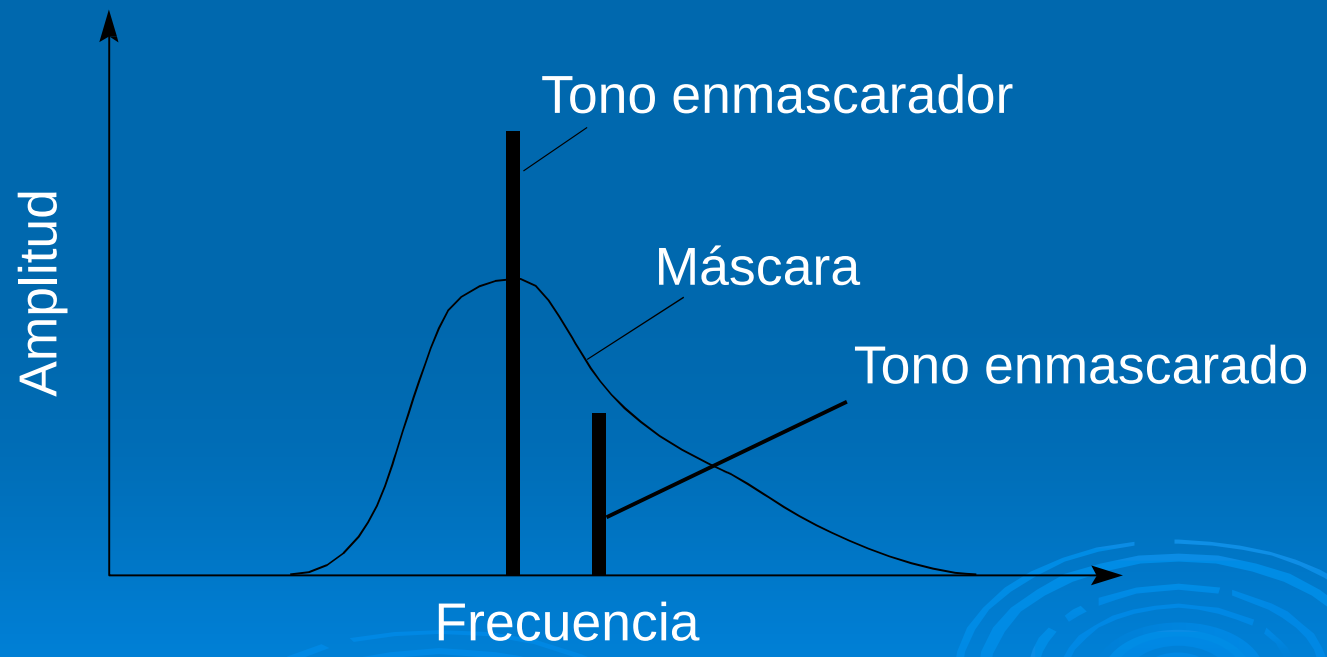
- No introducen distorsión
- Menor compresión ~ 1,5 a 6 veces
- Ejemplos: **FLAC** (formato libre y abierto), WavPack, Monkey's Audio, OptimFrog, LPAC, MPEG-4 ALS, Shorten, TAK, TTA.

Codificadores de audio perceptuales

Codificadores perceptuales

- Psicoacústica: El sistema auditivo humano funciona como un analizador de espectro de escala de bandas críticas (lineal de 20 a 500 Hz y logarítmica de 500 a 20000 Hz). Se aprovechan estas propiedades:
  - Enmascaramiento espectral: Los tonos más **intensos** enmascaran a los más **débiles** si están próximos en frecuencia.
  - Enmascaramiento temporal: Las señales más **intensas** enmascaran a las más **débiles**.
- El audio se comprime típicamente en un factor FC=10, pero la señal se distorsiona. En un oyente promedio, las distorsiones son apenas audibles o no audibles.

Enmascaramiento espectral



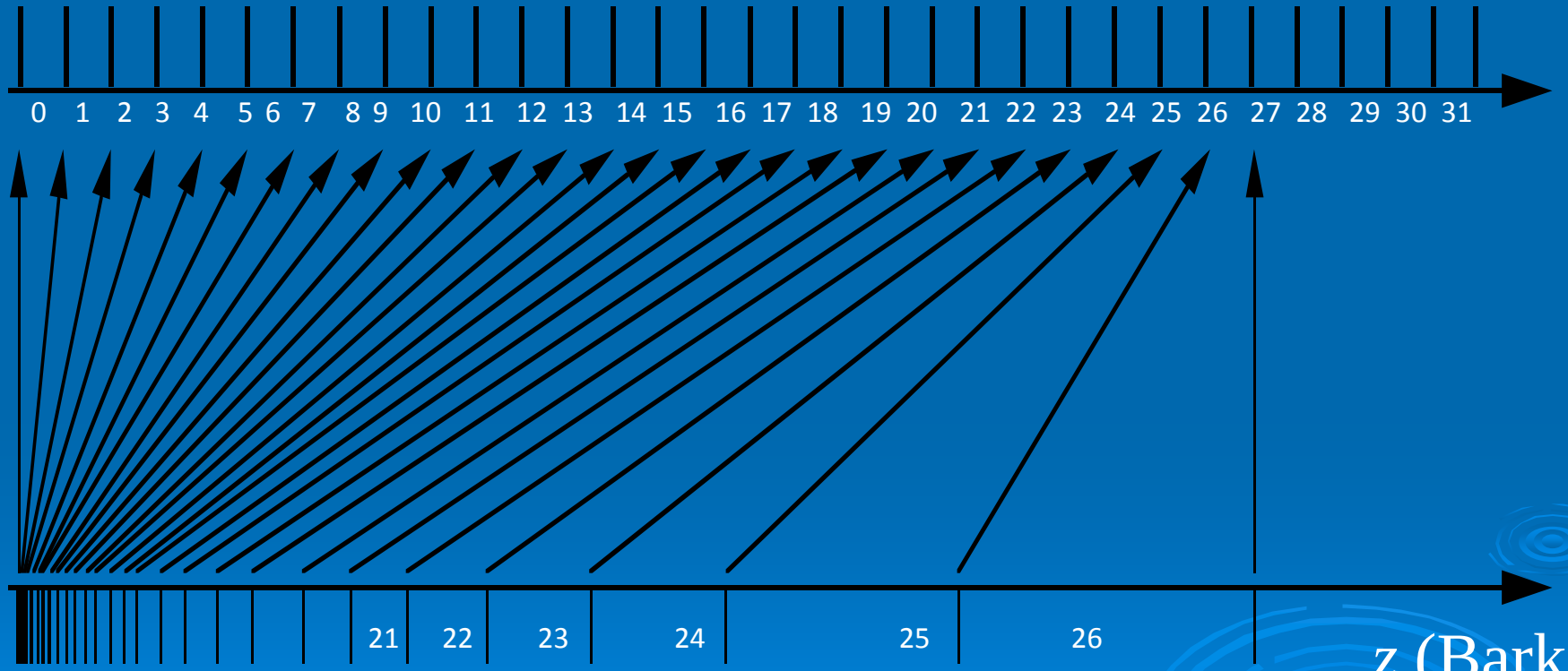
Bandas críticas

N°de banda (Bark)	Frecuencia (Hz) [1]	N°de banda (Bark)	Frecuencia (Hz) [1]
0	50	14	1970
1	95	15	2340
2	140	16	2720
3	235	17	3280
4	330	18	3840
5	420	19	4690
6	560	20	5440
7	660	21	6375
8	800	22	7690
9	940	23	9375
10	1125	24	11625
11	1265	25	15375
12	1500	26	20250
13	1735		

1 Valores de frecuencia en el límite superior de la banda.

Bandas críticas: mapeo

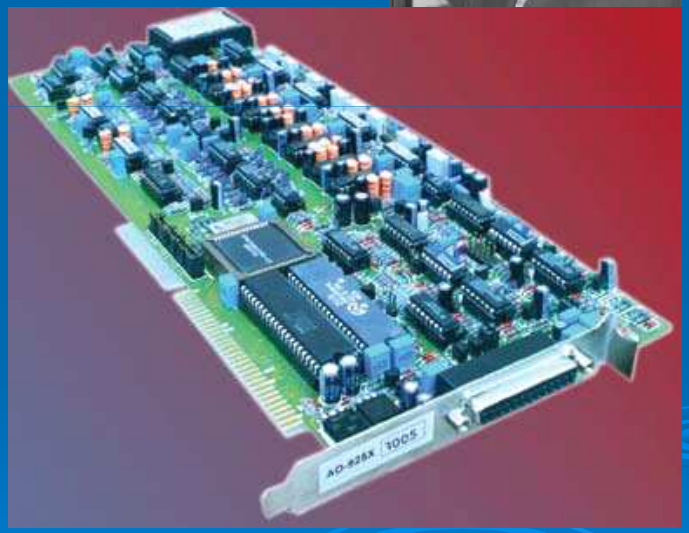
Frecuencias analizadas con banco de filtros



Ancho de bandas críticas

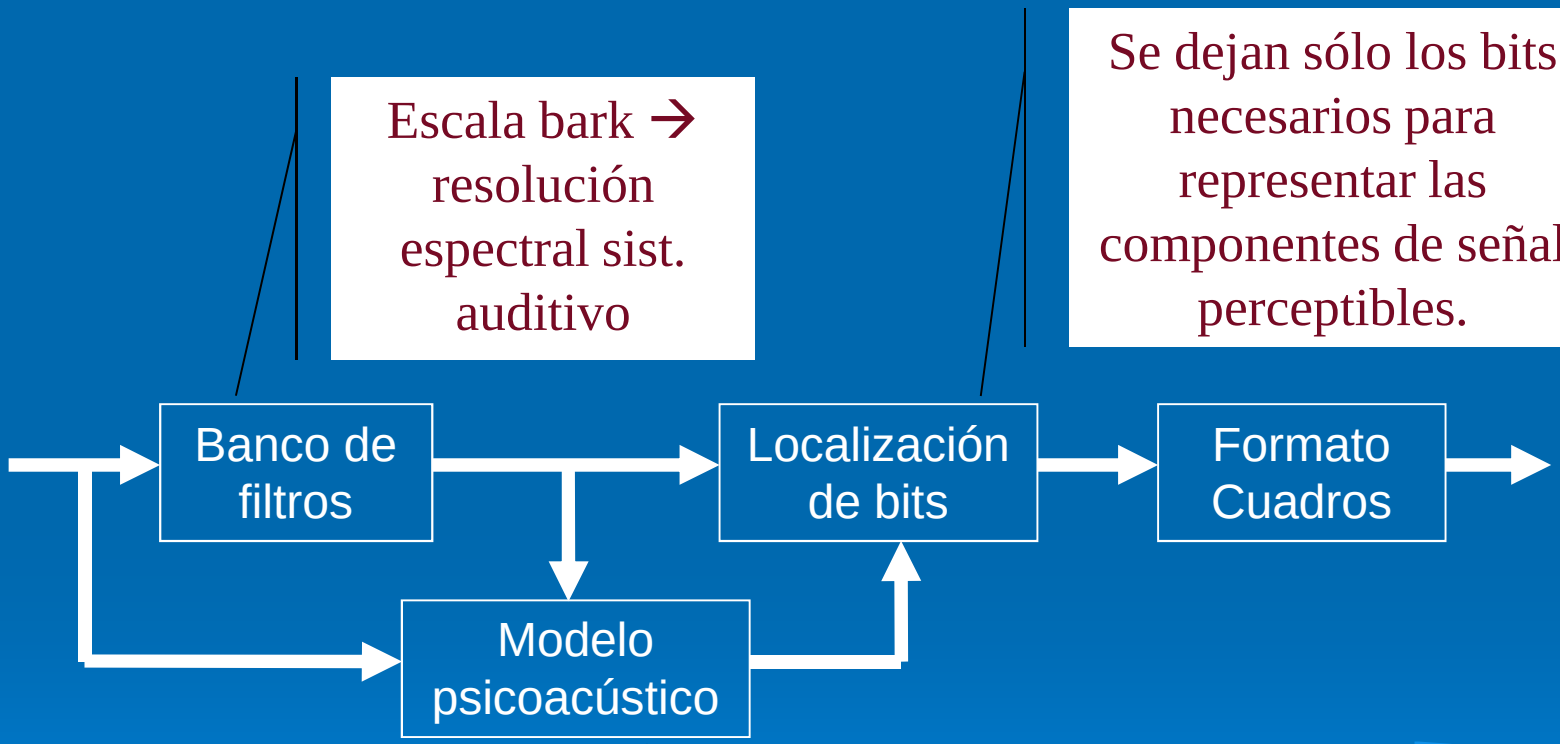
Codificadores perceptuales: pioneros

- En 1988, Oscar Bonello y su equipo diseñaron AUDICOM: la primer tecnología de compresión de audio perceptual por enmascaramiento (norma ECAM), para radiodifusión.
- Crearon una placa de audio enchufable en PC (no existente al momento) para comprimir y descomprimir audio.
- Crearon el software para ser manejado por un operador en tiempo real.

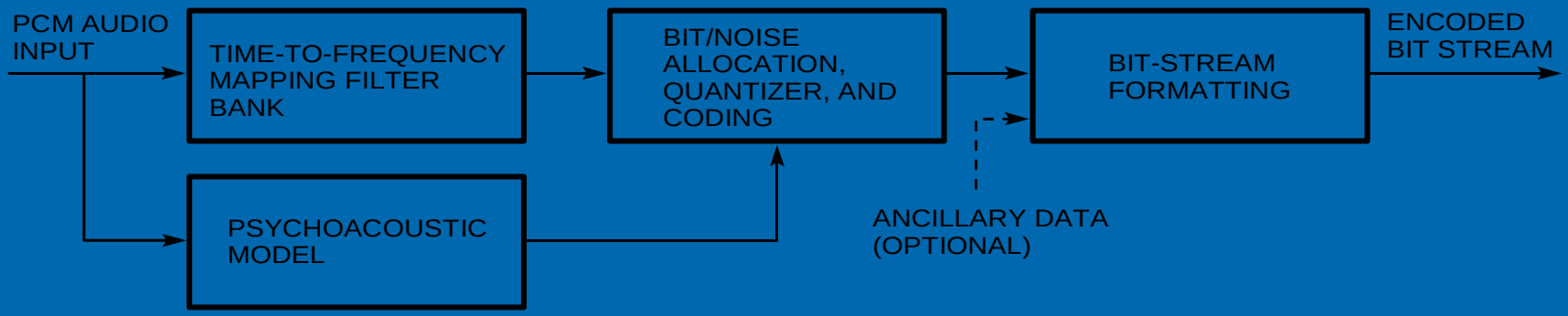


Bonello, O.J., *AUDICOM - Un invento argentino*, artículo en revista *Coordenadas*, año XXIV, N° 85, págs. 4-7, febrero-marzo 2010. Disponible en http://www.copitec.org.ar/comunicados/COORDENADAS_85.pdf

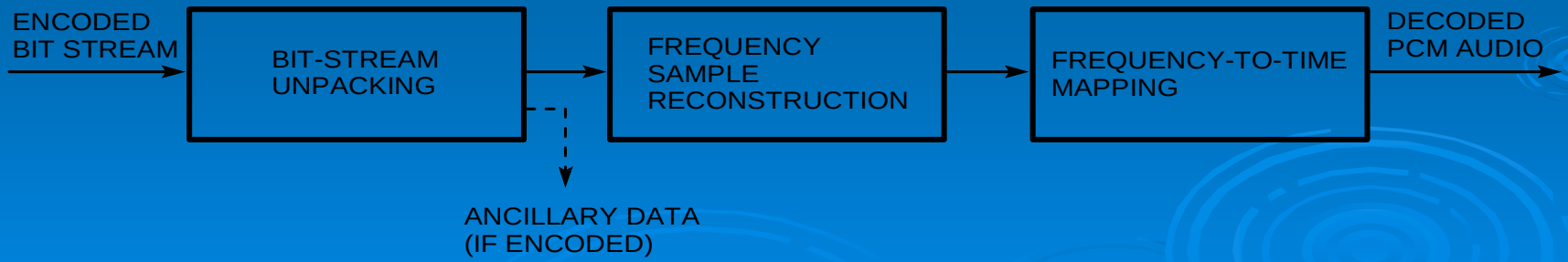
Codificadores perceptuales



Codificadores perceptuales: MP3



(a) MPEG/Audio Encoder



(b) MPEG/Audio Decoder

Otros codificadores perceptuales

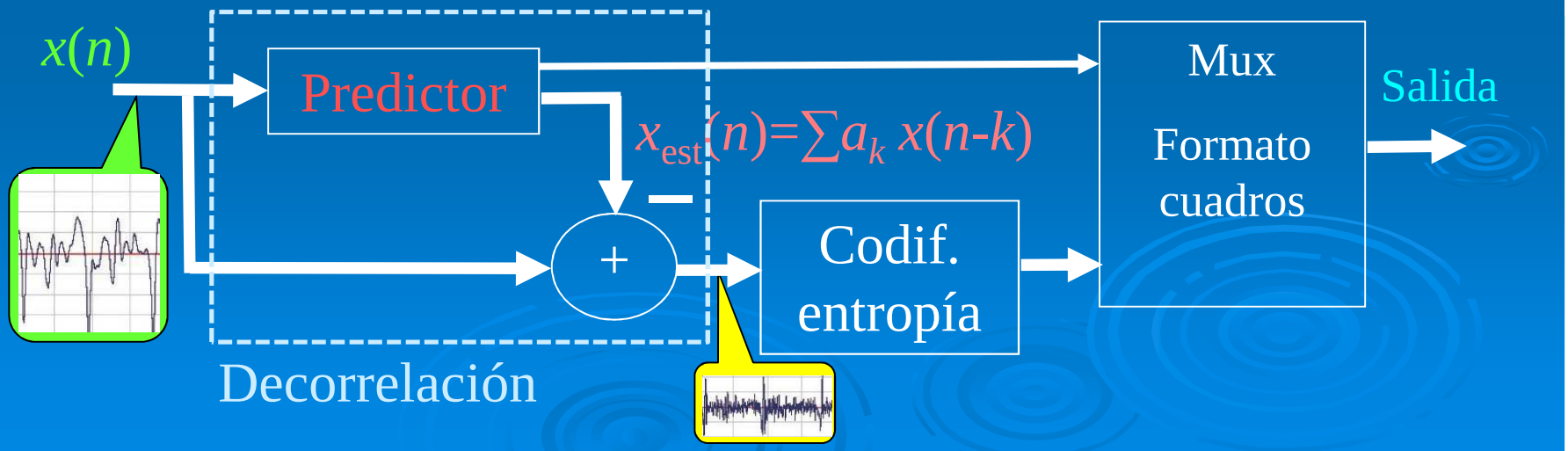
- Varios estándares: Advanced Audio Coding (AAC), AC3 (Dolby), etc.

Codificadores de audio sin pérdidas (CSP)

Codificación sin pérdidas

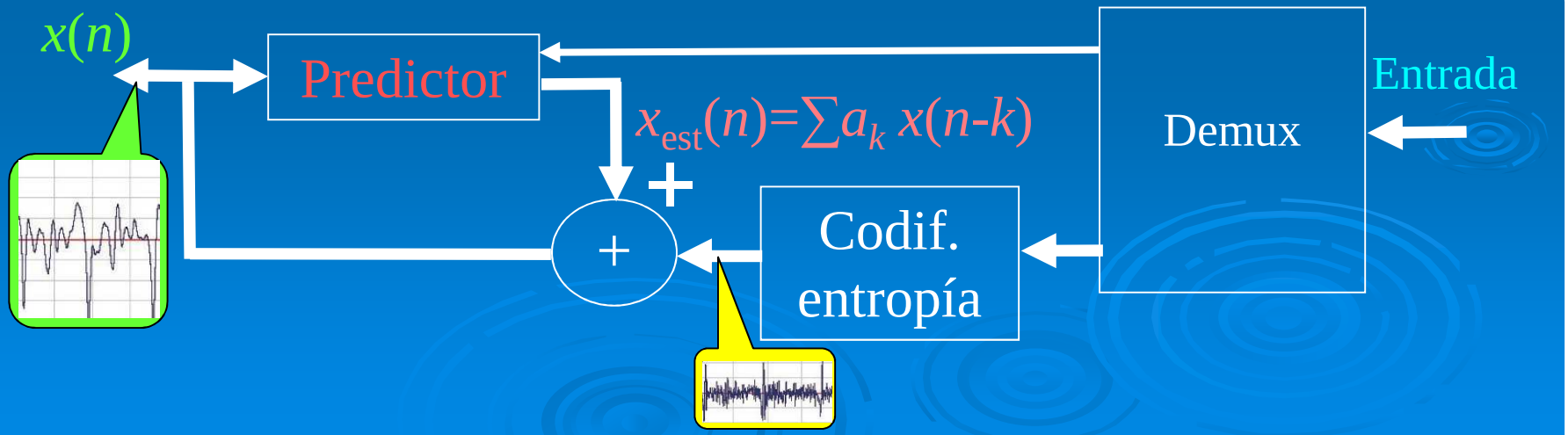
- **Predice** (aproxima) cada muestra de entrada usando valores pasados de entrada y quizás también de salida. Para esta predicción se utilizan muy pocos parámetros y el error de predicción (**residuo**) contiene mucha menos varianza (energía) que la señal de entrada

  → necesito menos bits para representar el residuo



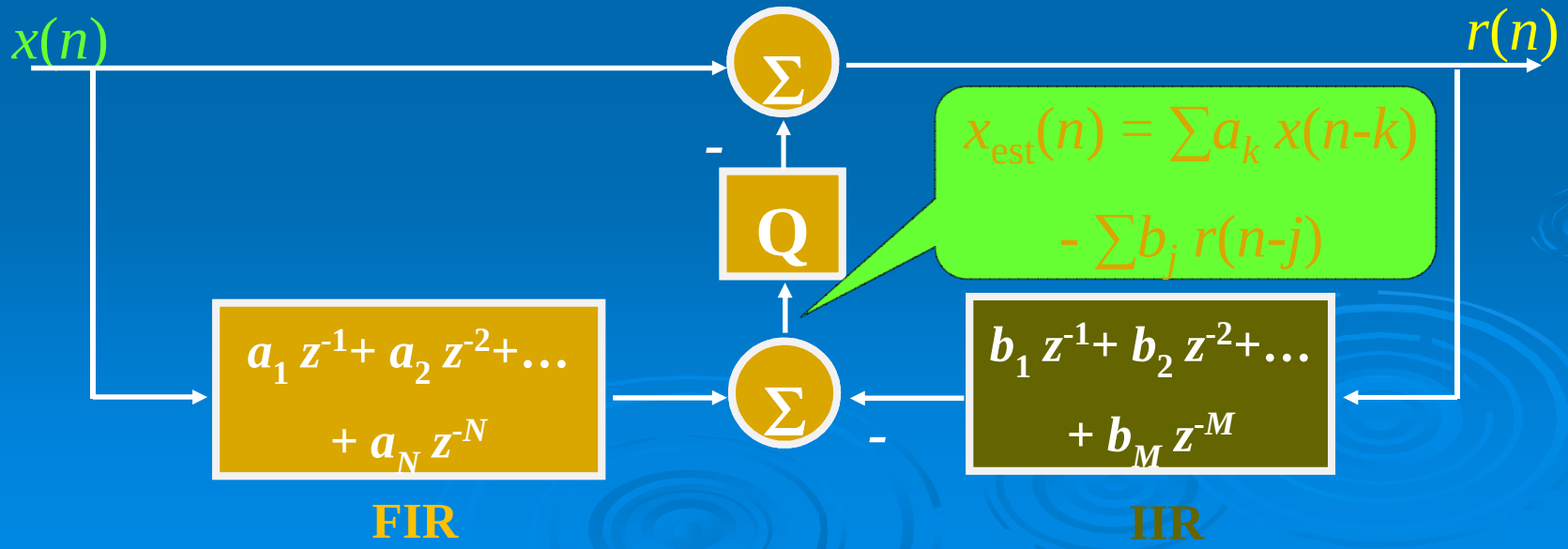
Decodificación sin pérdidas

- Se utilizan los mismos parámetros del predictor que en el codificador: se demultiplexan los **datos comprimidos**, se decodifica el **residuo** y se suma a éste el valor actual de **salida predicha**.



Codificación sin pérdidas (CSP)

- Caso típico: predictor FIR. Generalmente es un modelo de coeficientes de predicción lineal (LPC). Estos coeficientes se ajustan para minimizar la energía del error de predicción.
- En pocos casos se utilizan coeficientes IIR.



¿Por qué están en auge los CSP?

- **Tecnología:**
  - Recursos de almacenamiento con mayor capacidad de memoria.
  - La creciente velocidad de transmisión de datos vía Internet minimiza la ventaja de los codec perceptuales frente a los CSP.
- **Razones económicas:**
  - En los equipos de reproducción de sonido de alta fidelidad se vuelve evidente la distorsión que introduce la codificación perceptual. Dichos equipos disminuyen de costo y su acceso es más masivo.
- **Masterización:**
  - Las grabaciones de estudio se pueden transmitir entre equipos remotos en formato comprimido sin distorsión: la única opción viable es usar CSP.

Nuevos codecs de audio basados en EMD

Algoritmo de EMD

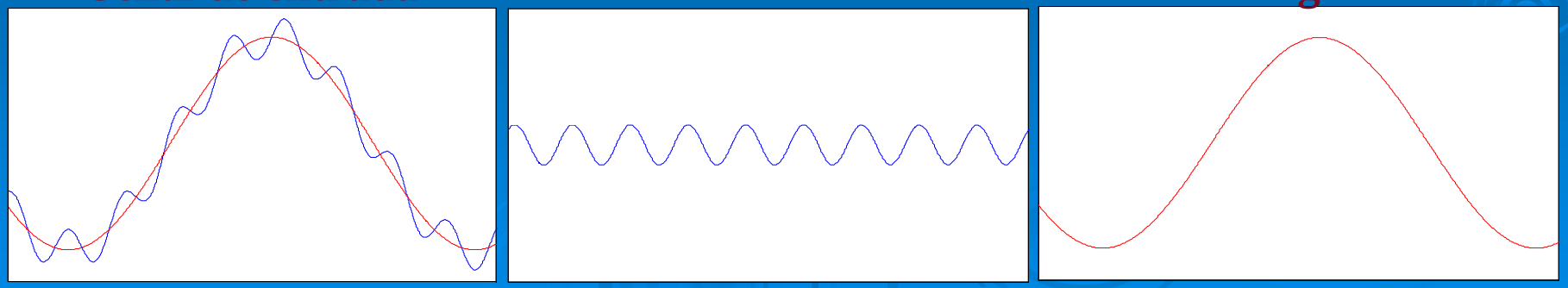
- Descompone una señal 1D en una suma de señales (IMF) con distinto nivel de resolución empezando por el detalle más fino.
- Es adaptativo y depende sólo de las componentes existentes en la señal analizada.
- **Función de modo intrínseco ó IMF:** es de tipo AM-FM de banda limitada, tiene igual cantidad de extremos que de cruces por cero y la media entre las envolventes superior e inferior es nula.

Huang et al. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". Proc. Royal Soc. of London, Ser. A, 454, pp.903–995.

Algoritmo de EMD

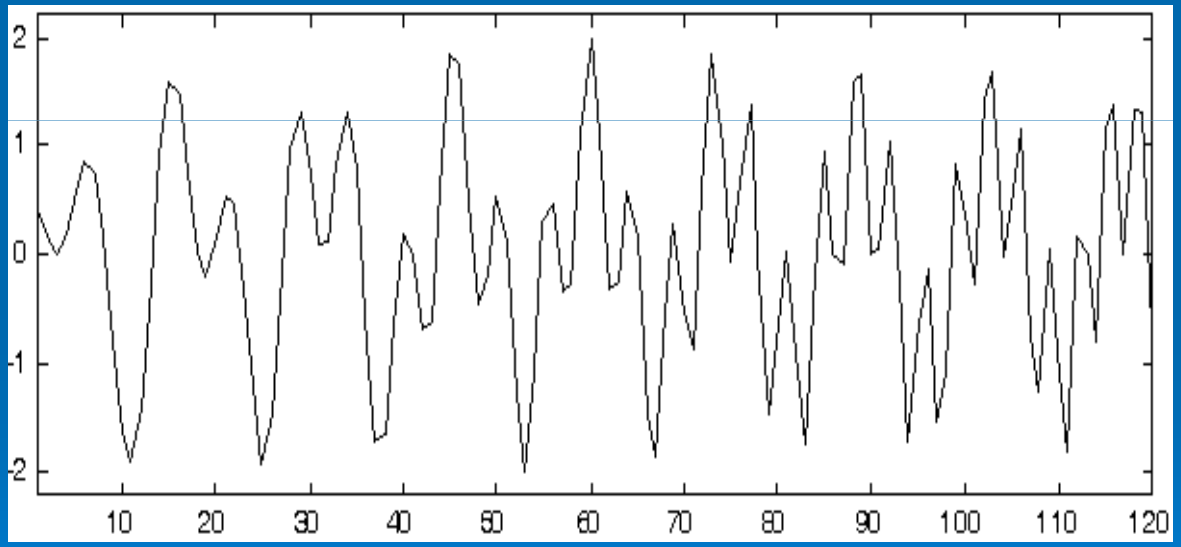
- Analogía: Fourier descompone la señal en funciones de frecuencia y amplitud fijas, EMD descompone en funciones oscilatorias de frecuencia y amplitud variables.
- Ventaja: Las funciones IMF son monocomponentes (única frecuencia y amplitud en cada instante)
  - → la teoría permite calcular dicha frecuencia y amplitud
  - → se minimiza la influencia del principio de incertidumbre tiempo-frecuencia.

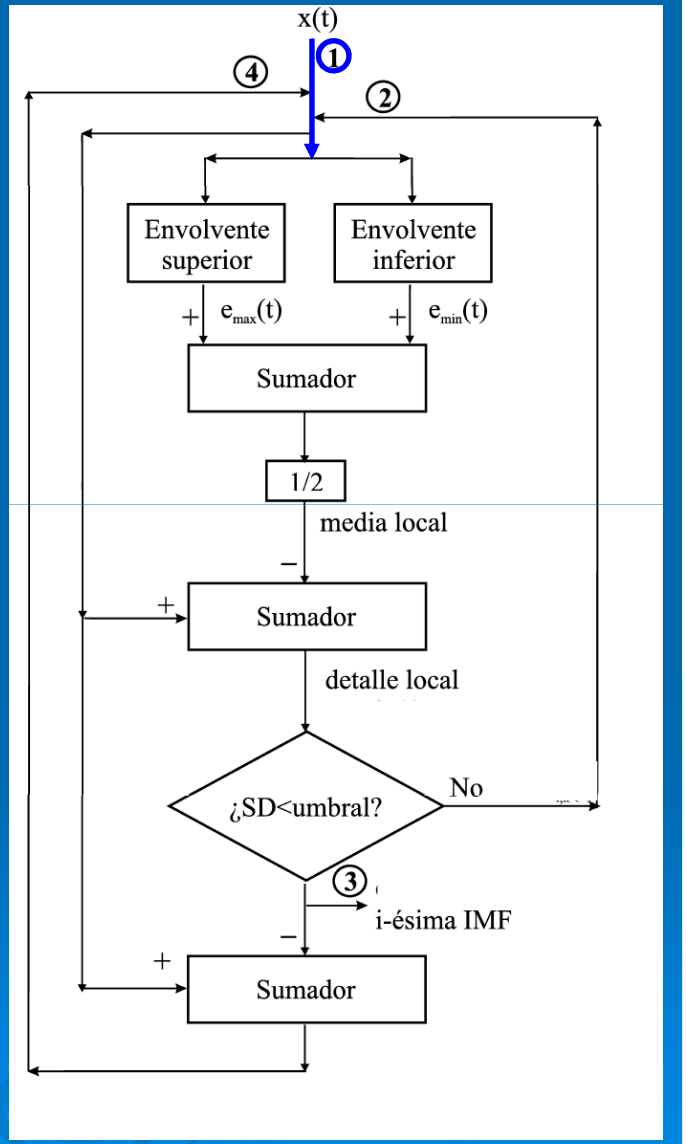
Señal de entrada **Detalle fino** **Detalle grueso**

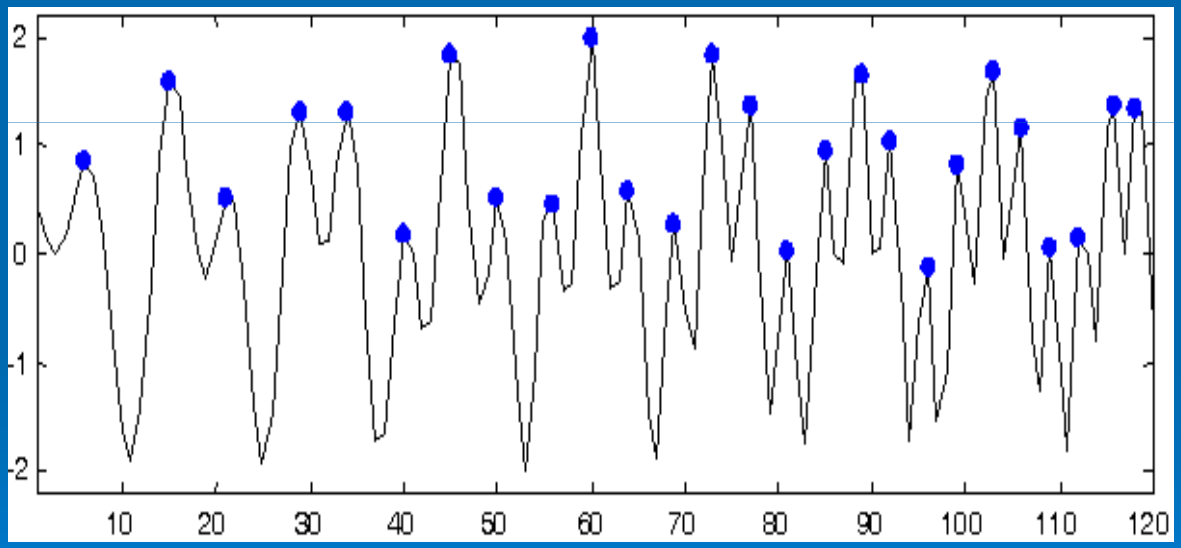


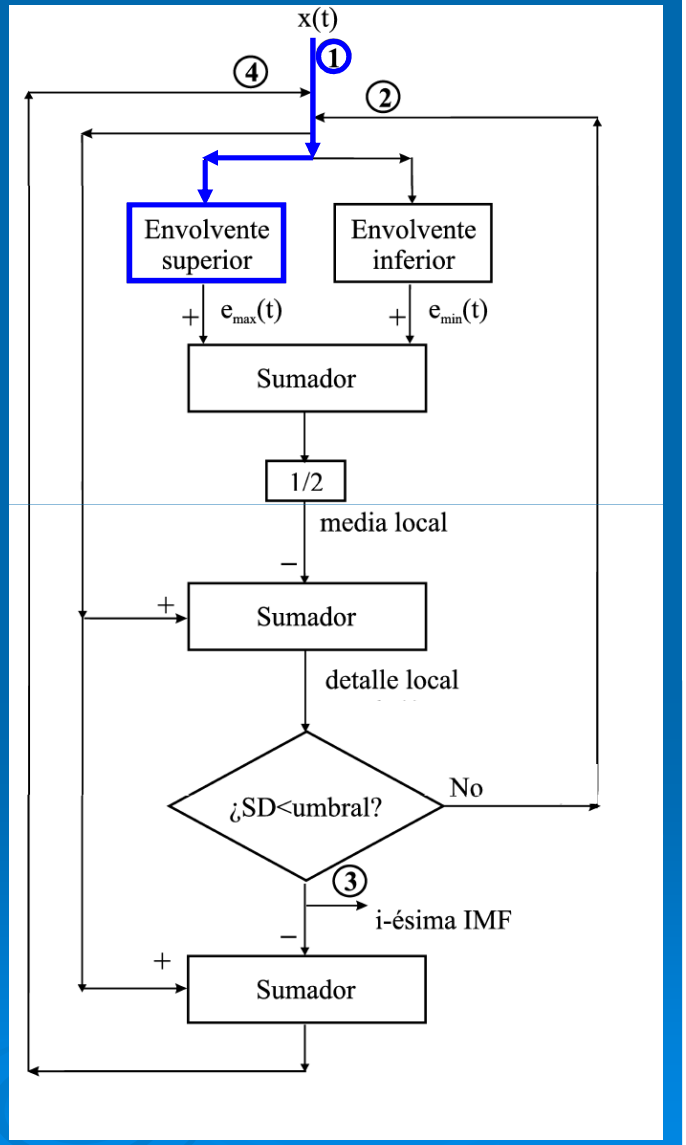
Algoritmo de EMD: tamizado

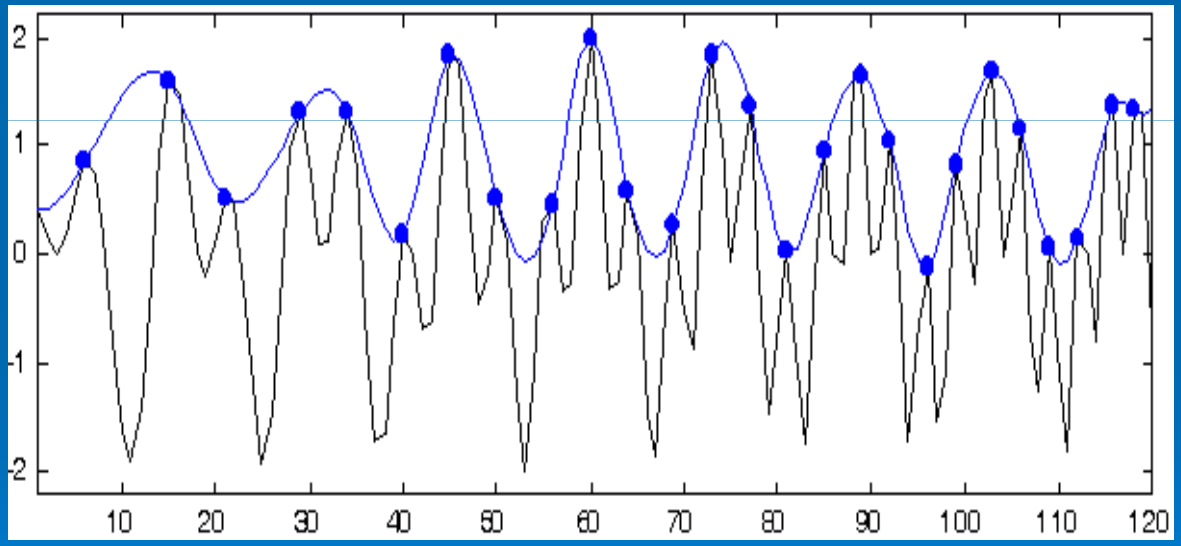
1) Señal original: $x(t)$

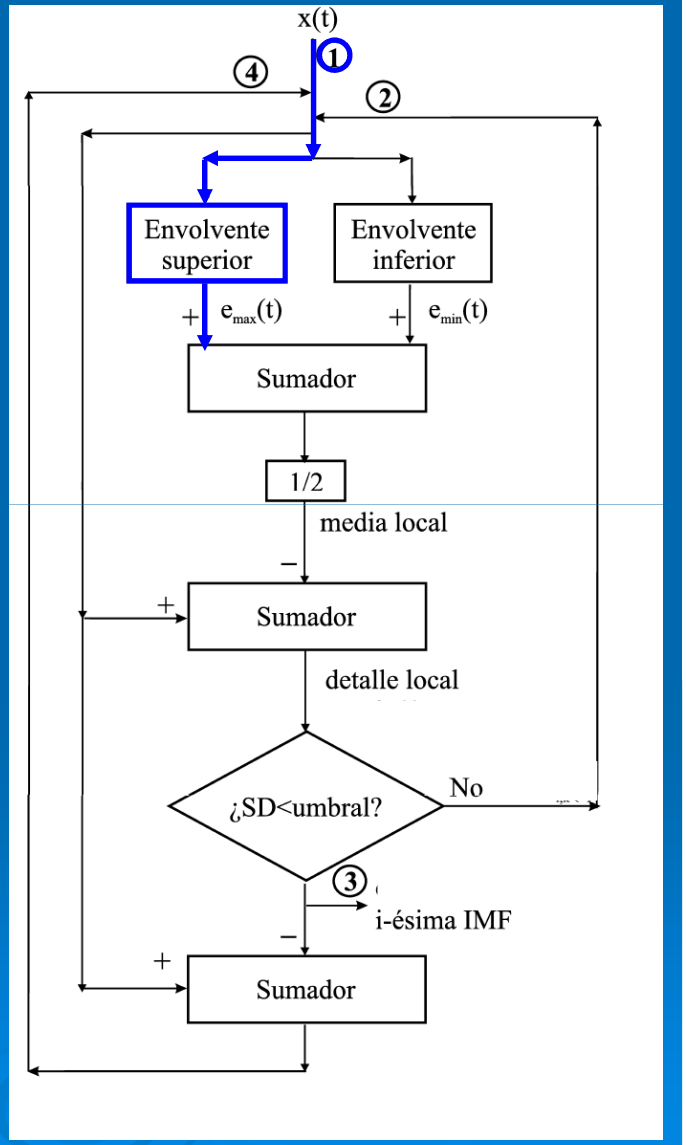


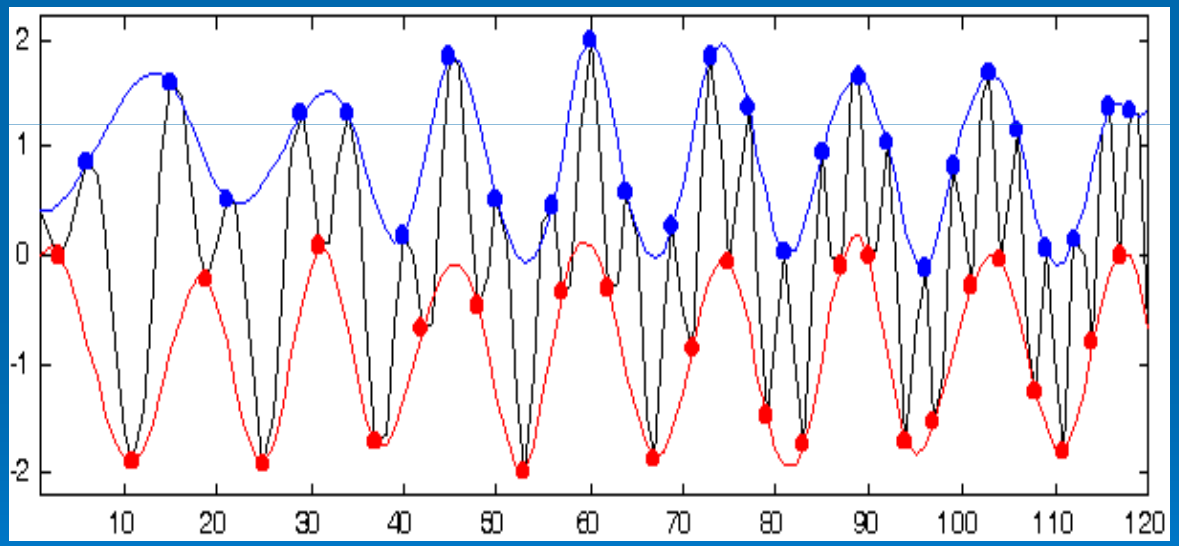


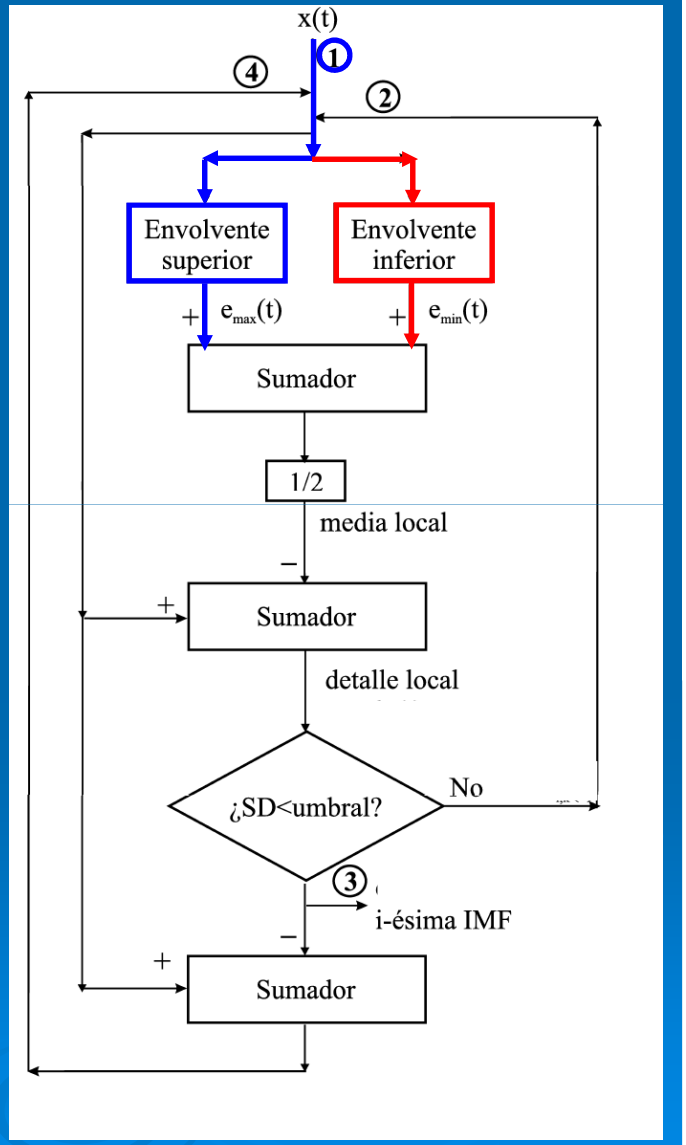

2
1
0
-1
-2
10
20
30
40
50
60
70
80
90
100
110
120


x(t)
1
2
3
4
Envolvente superior
Envolvente inferior
+
$e_{max}(t)$
+
$e_{min}(t)$
Sumador
1/2
media local
−
+
Sumador
detalle local
¿SD<umbral?
No
i-ésima IMF
−
+
Sumador

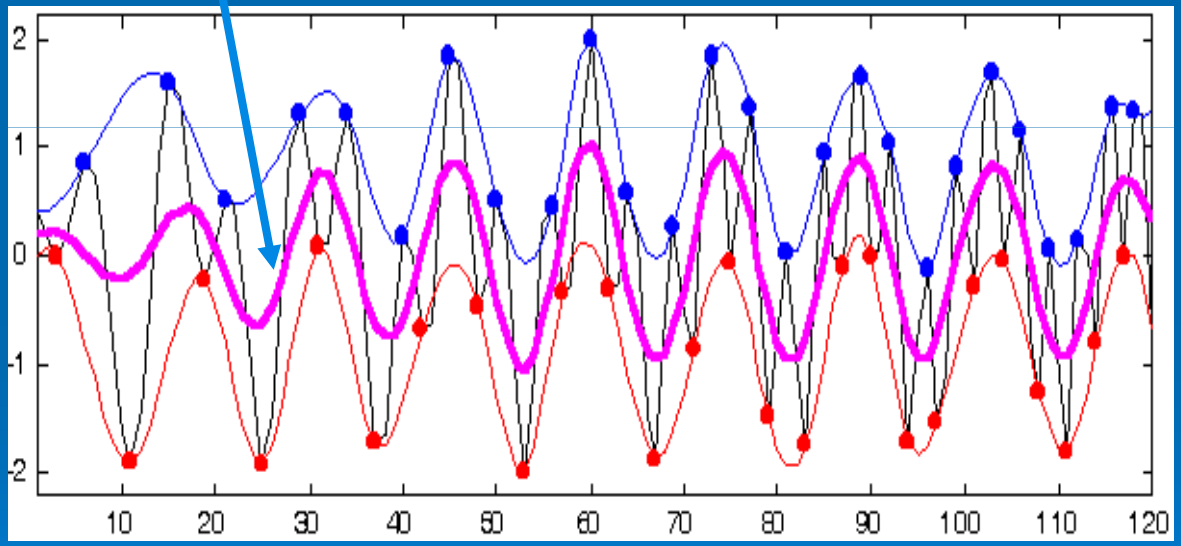

2
1
0
-1
-2
10
20
30
40
50
60
70
80
90
100
110
120


x(t)
1
2
3
4
Envolvente superior
Envolvente inferior
+
$e_{max}(t)$
+
$e_{min}(t)$
Sumador
1/2
media local
−
+
Sumador
detalle local
¿SD<umbral?
No
i-ésima IMF
−
+
Sumador

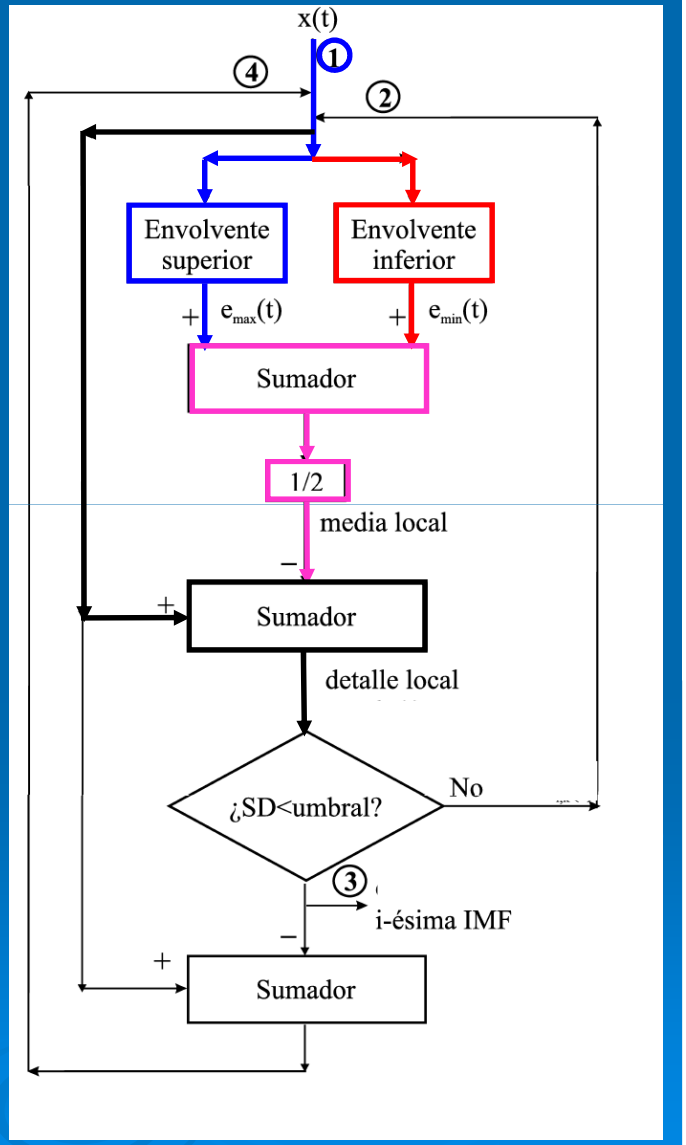

2
1
0
-1
-2
10
20
30
40
50
60
70
80
90
100
110
120

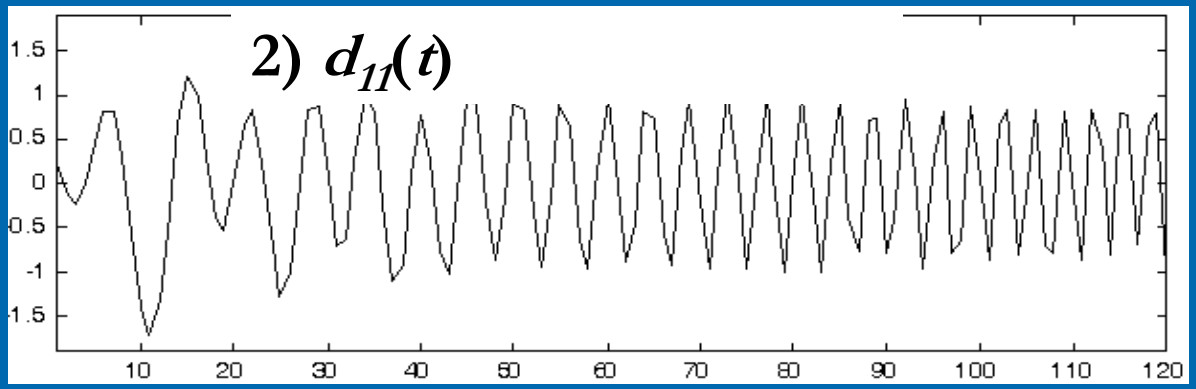

x(t)
1
2
3
4
Envolvente superior
Envolvente inferior
+
$e_{max}(t)$
+
$e_{min}(t)$
Sumador
1/2
media local
−
+
Sumador
detalle local
¿SD<umbral?
No
i-ésima IMF
−
+
Sumador

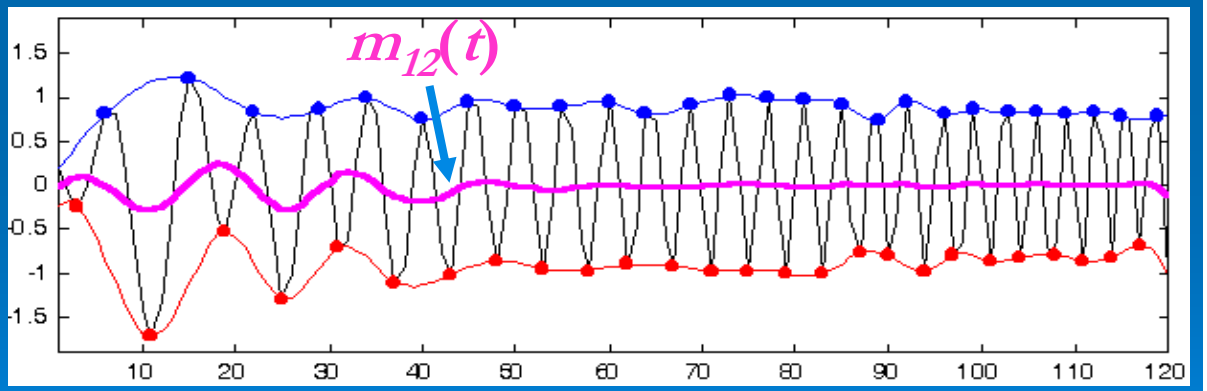
$m_{11}(t)$



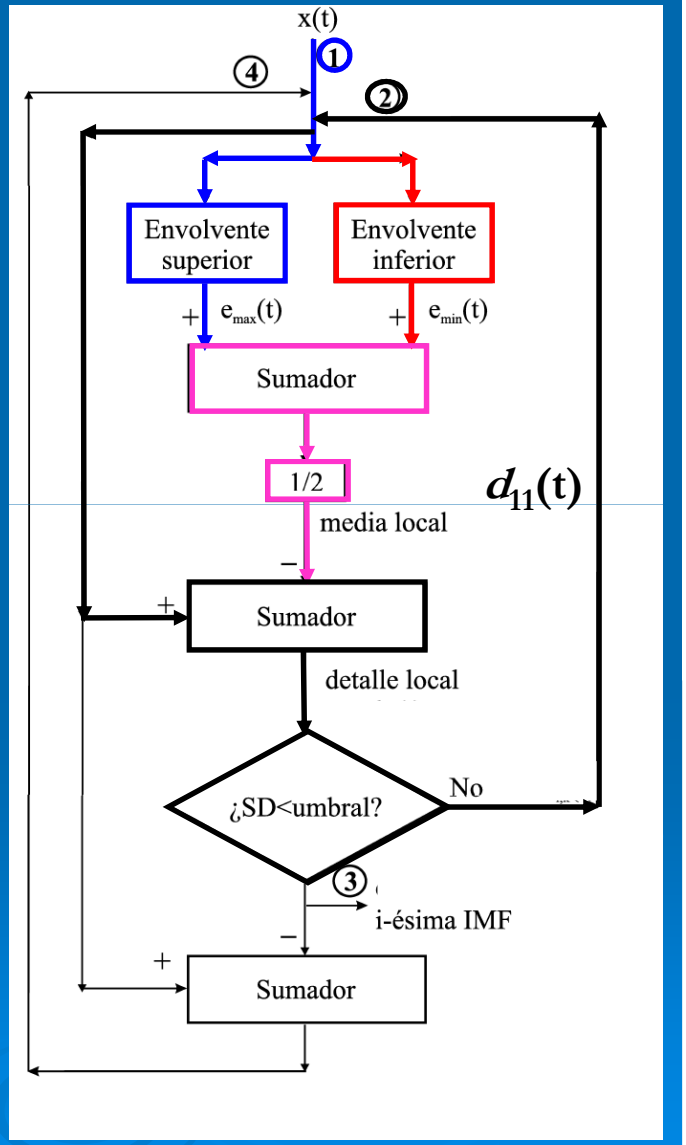
$$d_{11}(t) = x(t) - m_{11}(t)$$





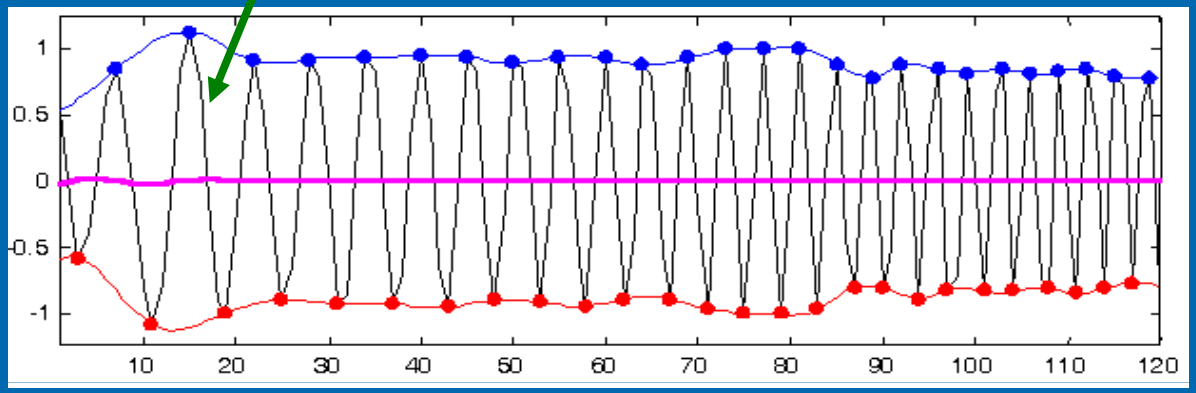


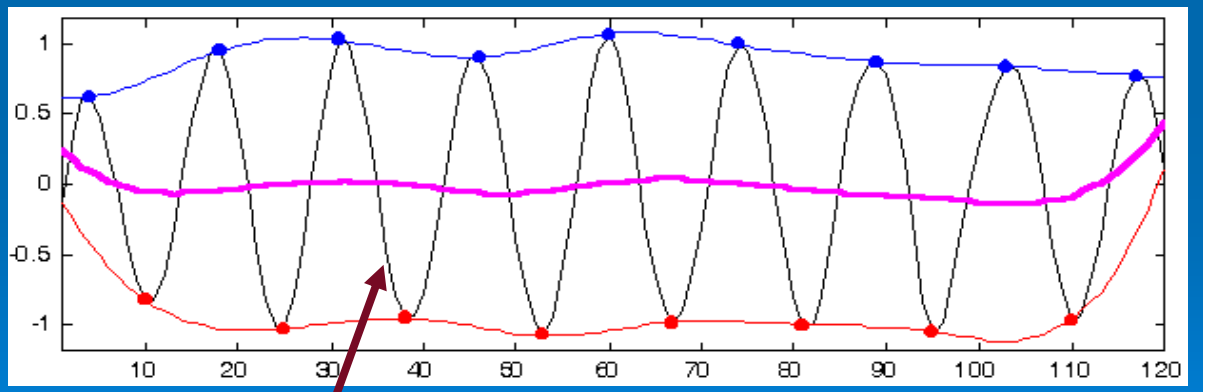
$$d_{12} = d_{11} - m_{12} \ldots d_{1k} = d_{1(k-1)} - m_{1k}$$



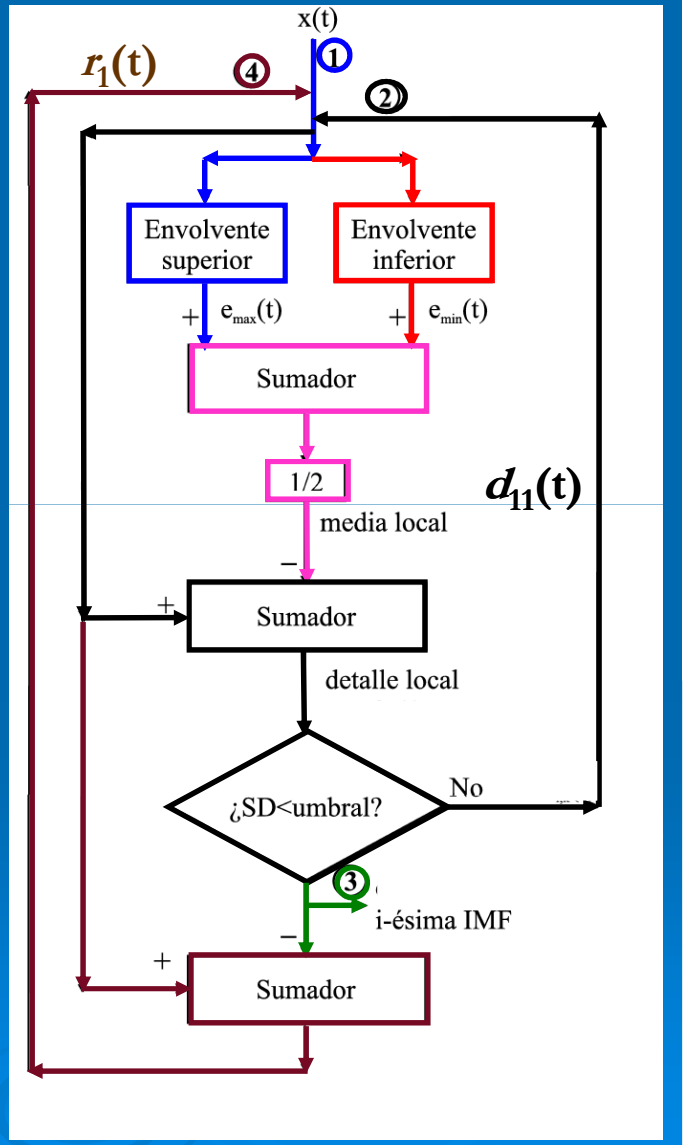
Tamizado: primera IMF

3) Primera IMF: $c_1 = d_{1k}$





4) Primer residuo $r_1(t)$



Fin del tamizado

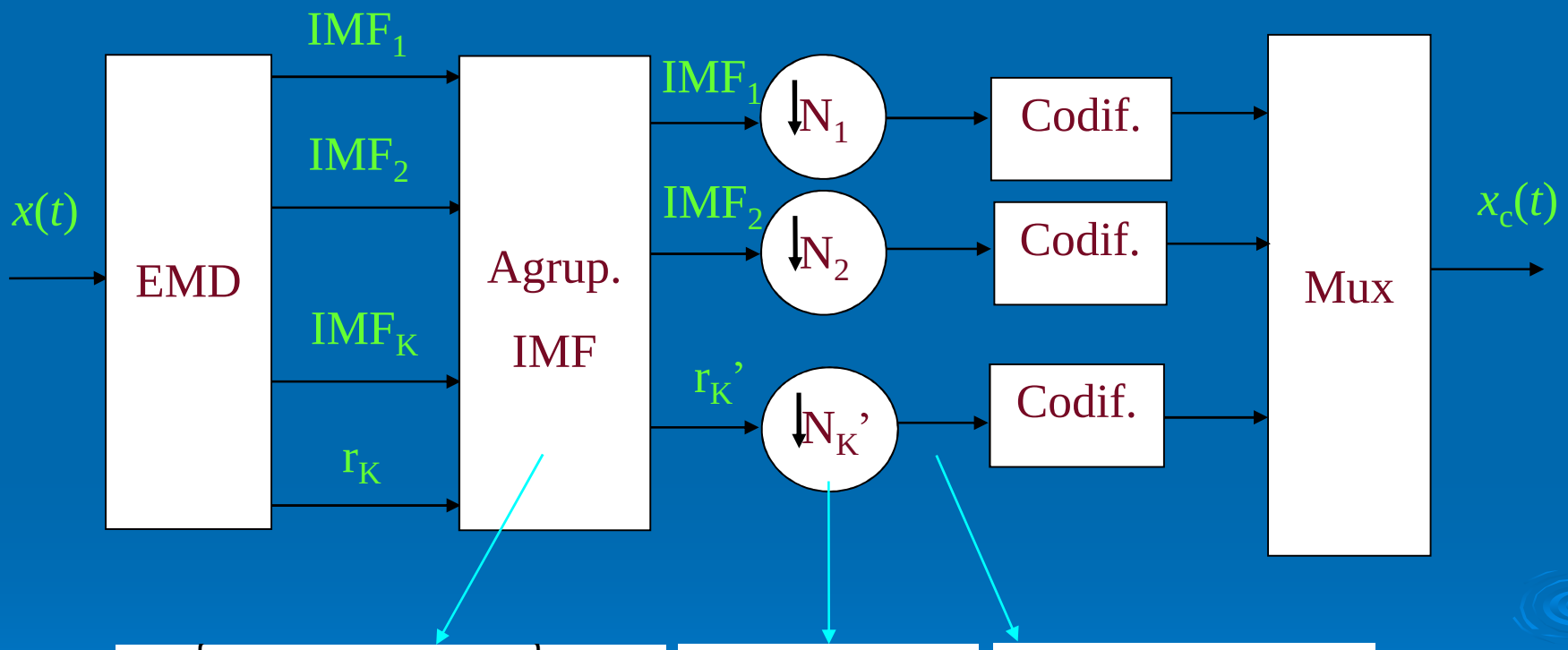
- El tamizado finaliza cuando no quedan más oscilaciones para extraer en el residuo, o éste tiene solo un máximo y un mínimo local
  → residuo final
- La señal de entrada se expresa como

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t)$$

IMF

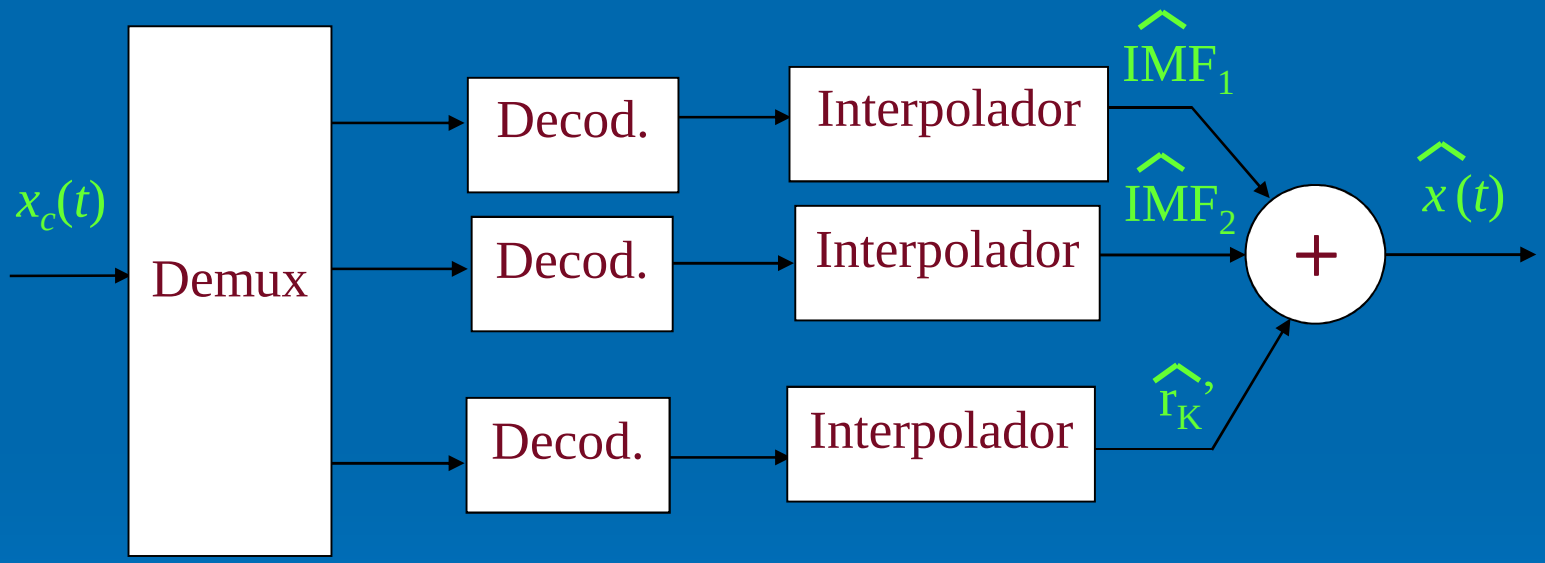
Residuo final

Codificador EMD propuesto

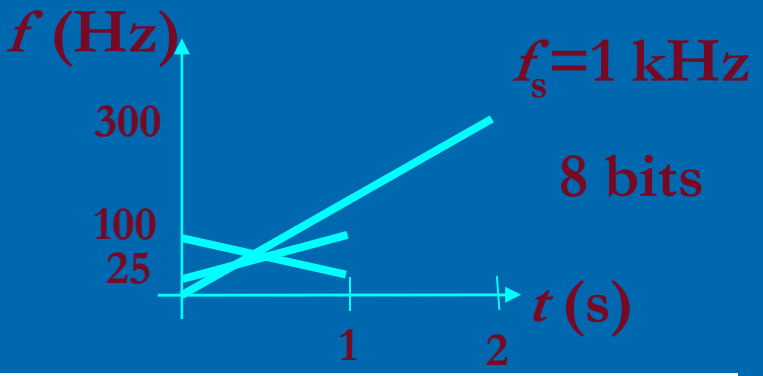


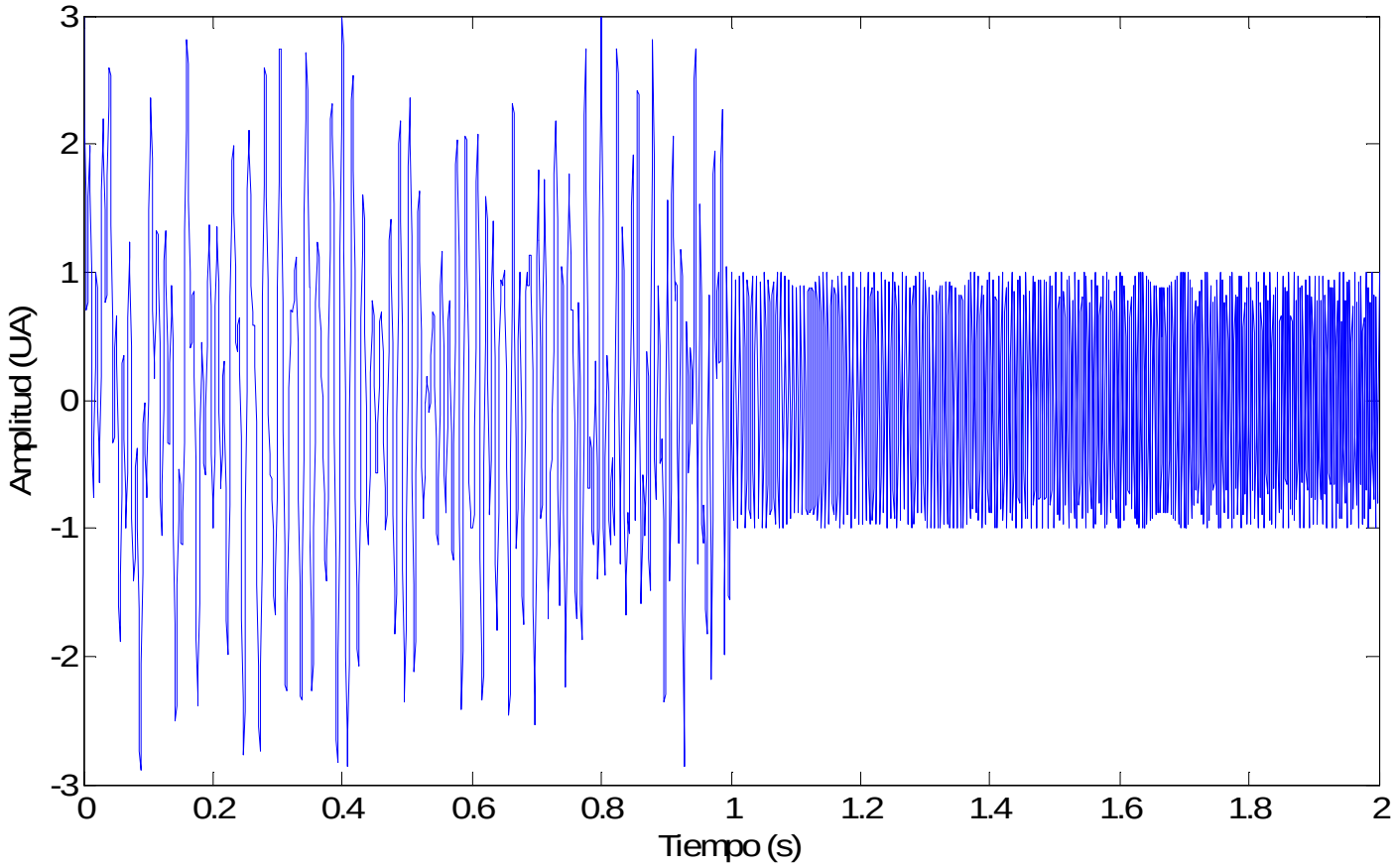
Marengo Rodriguez, Miyara. “Representación de señales de audio con descomposición empírica de modos y submuestreo adaptativo”, AdAA 2009, A056 Rosario.

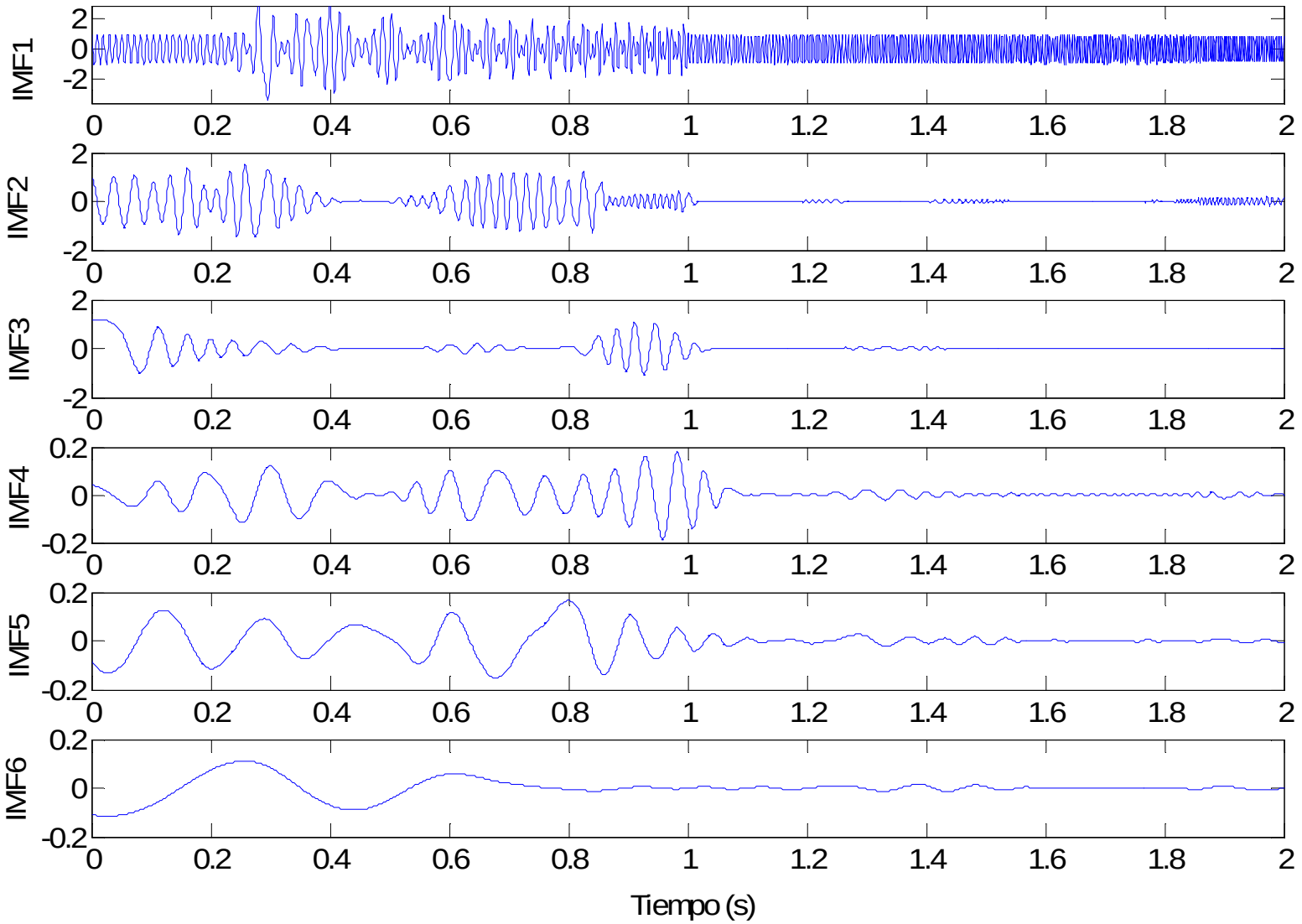
Decodificador EMD propuesto

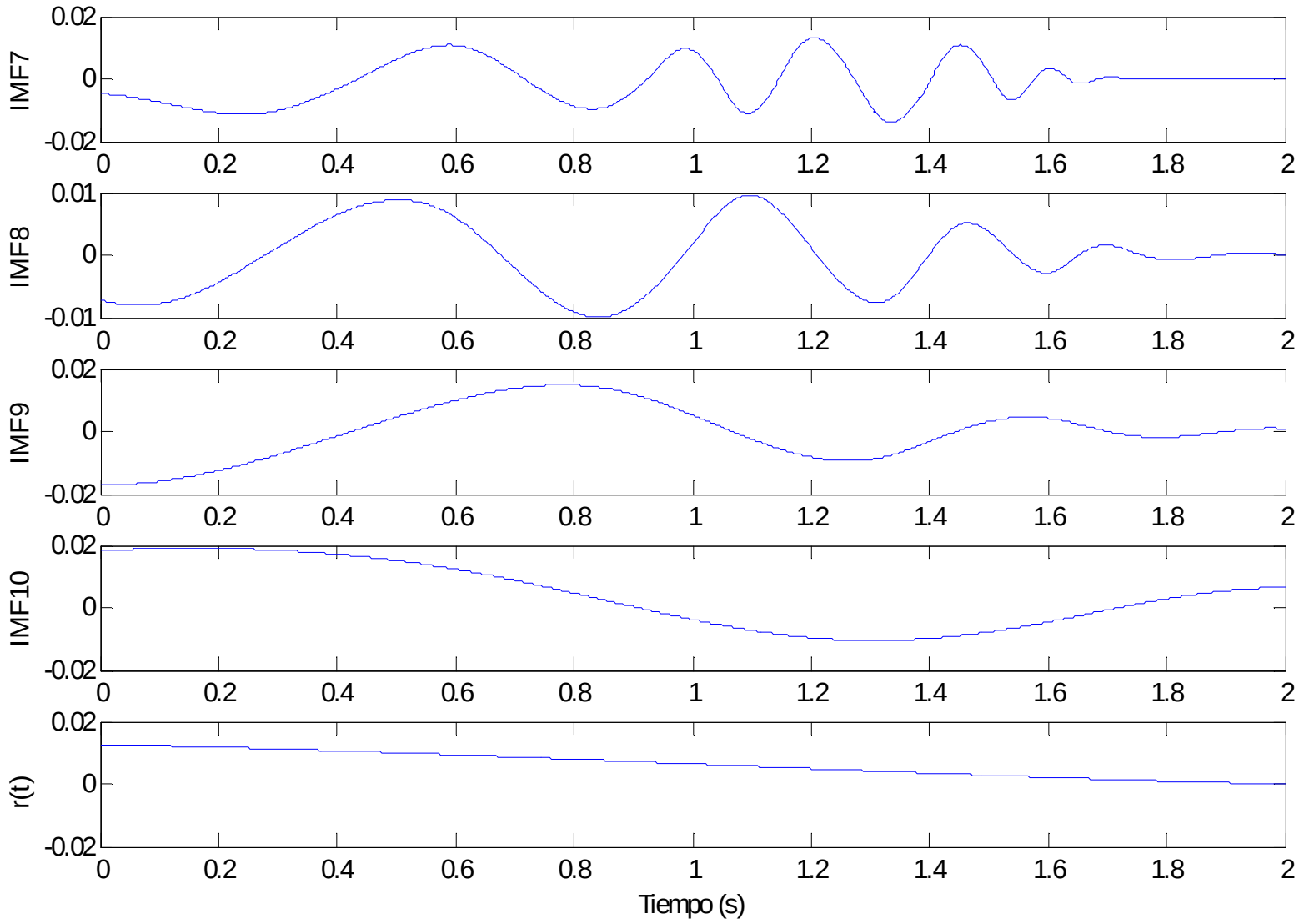


Ejemplo: Σ chirp

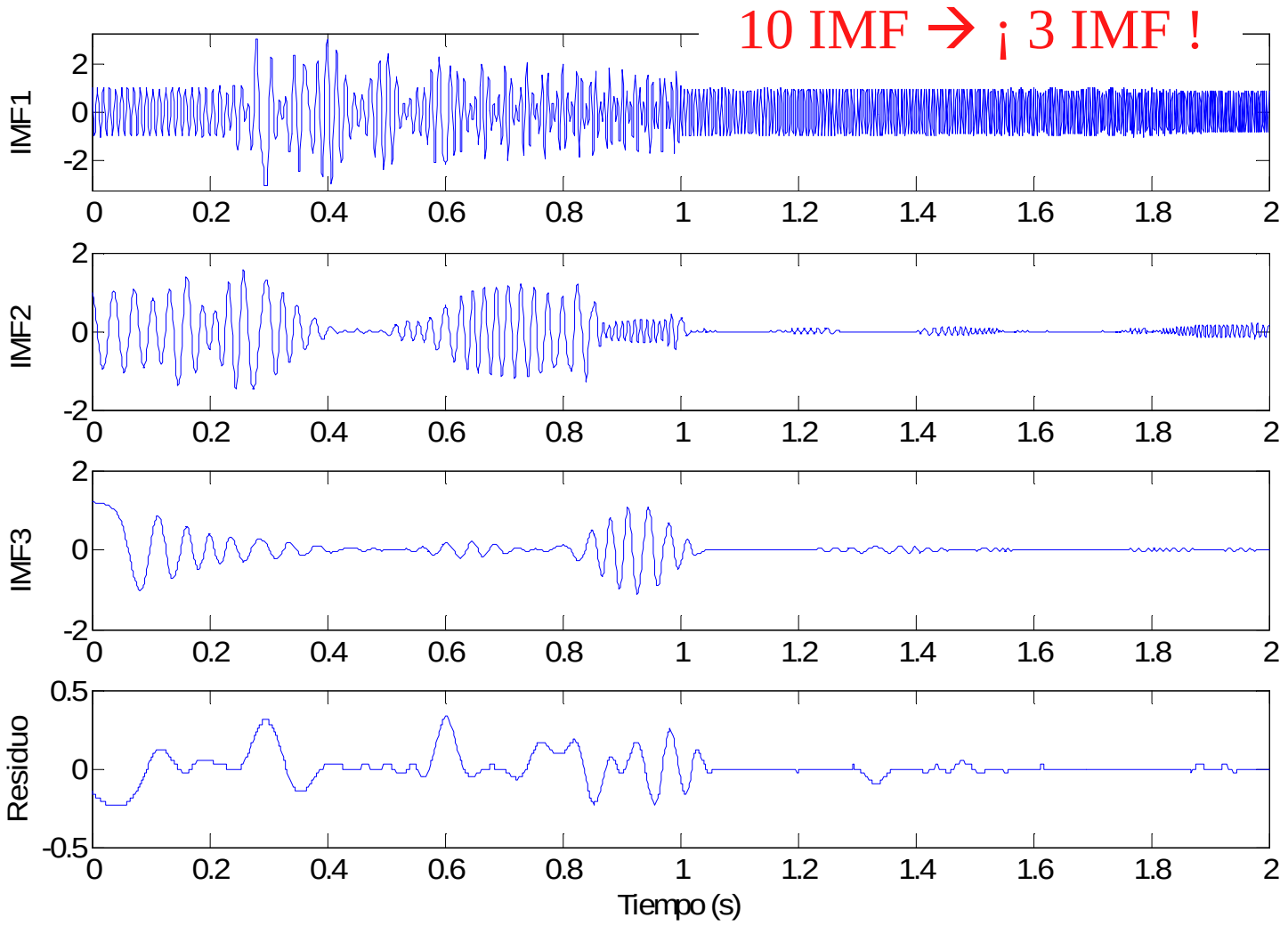




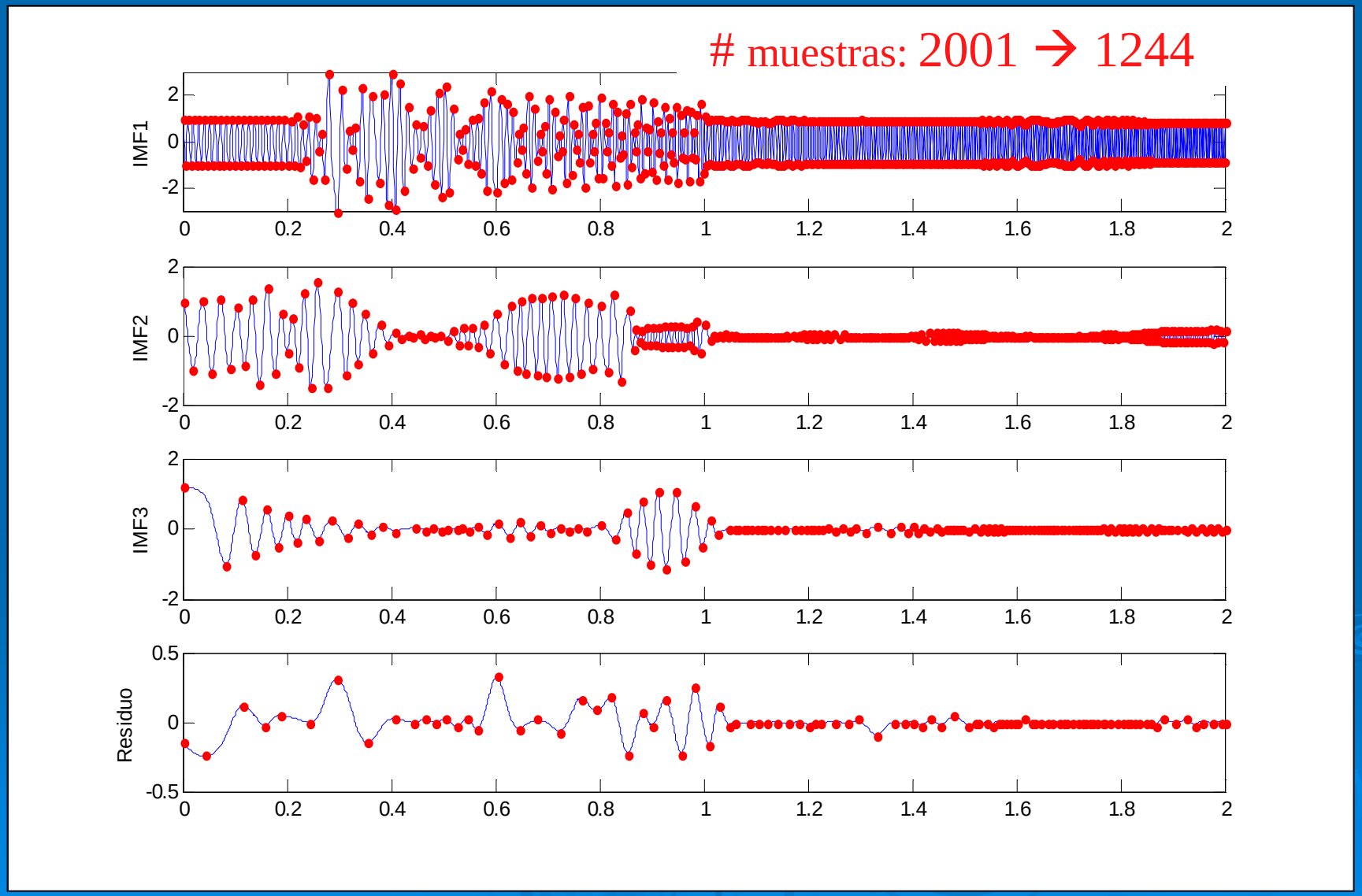

IMF1
IMF2
IMF3
IMF4
IMF5
IMF6
2
0
-2
0.2
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Tiempo (s)


IMF7
0.02
0
-0.02
IMF8
0.01
0
-0.01
IMF9
0.02
0
-0.02
IMF10
0.02
0
-0.02
r(t)
0.02
0
-0.02
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
Tiempo (s)

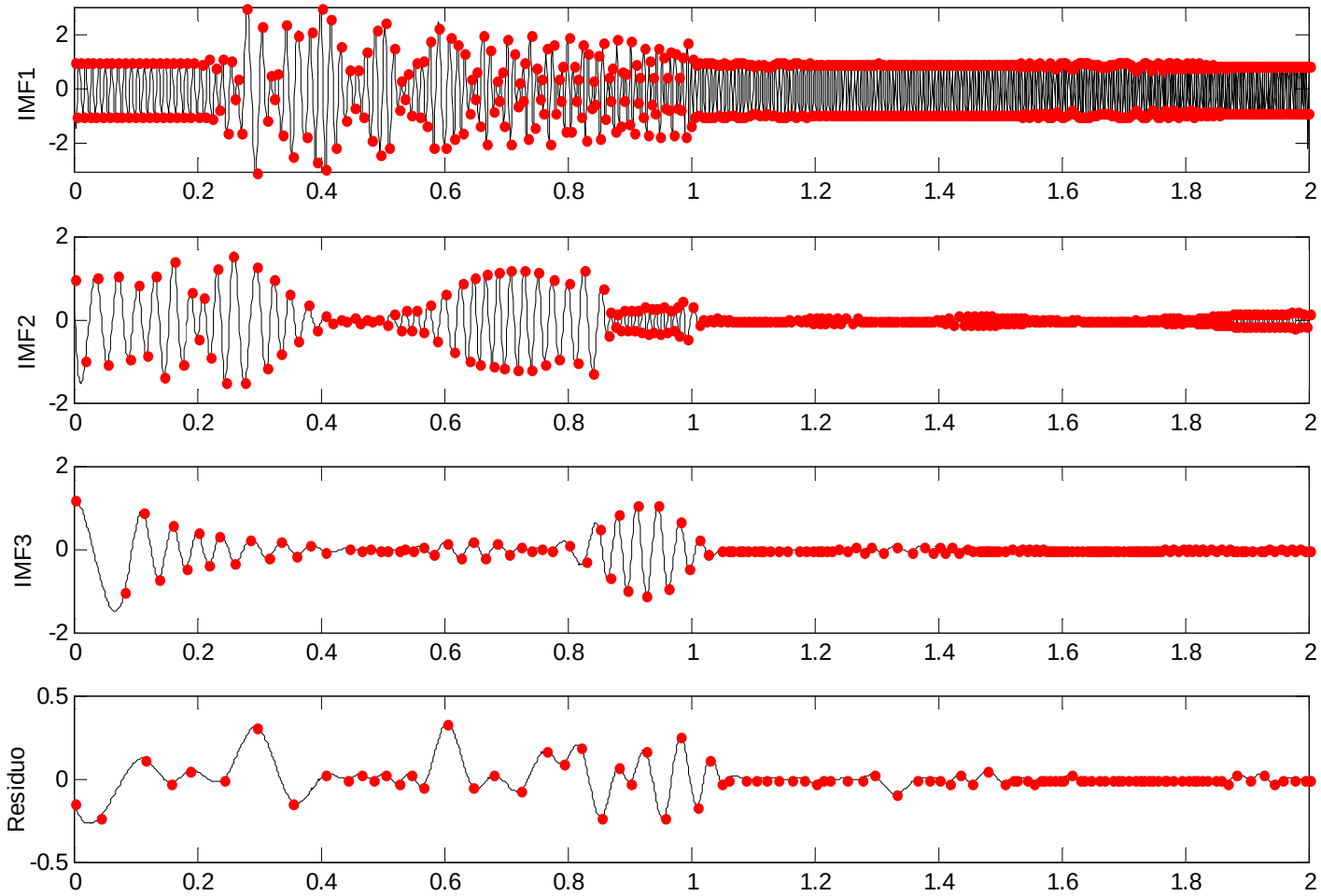
Agrupamiento de IMF



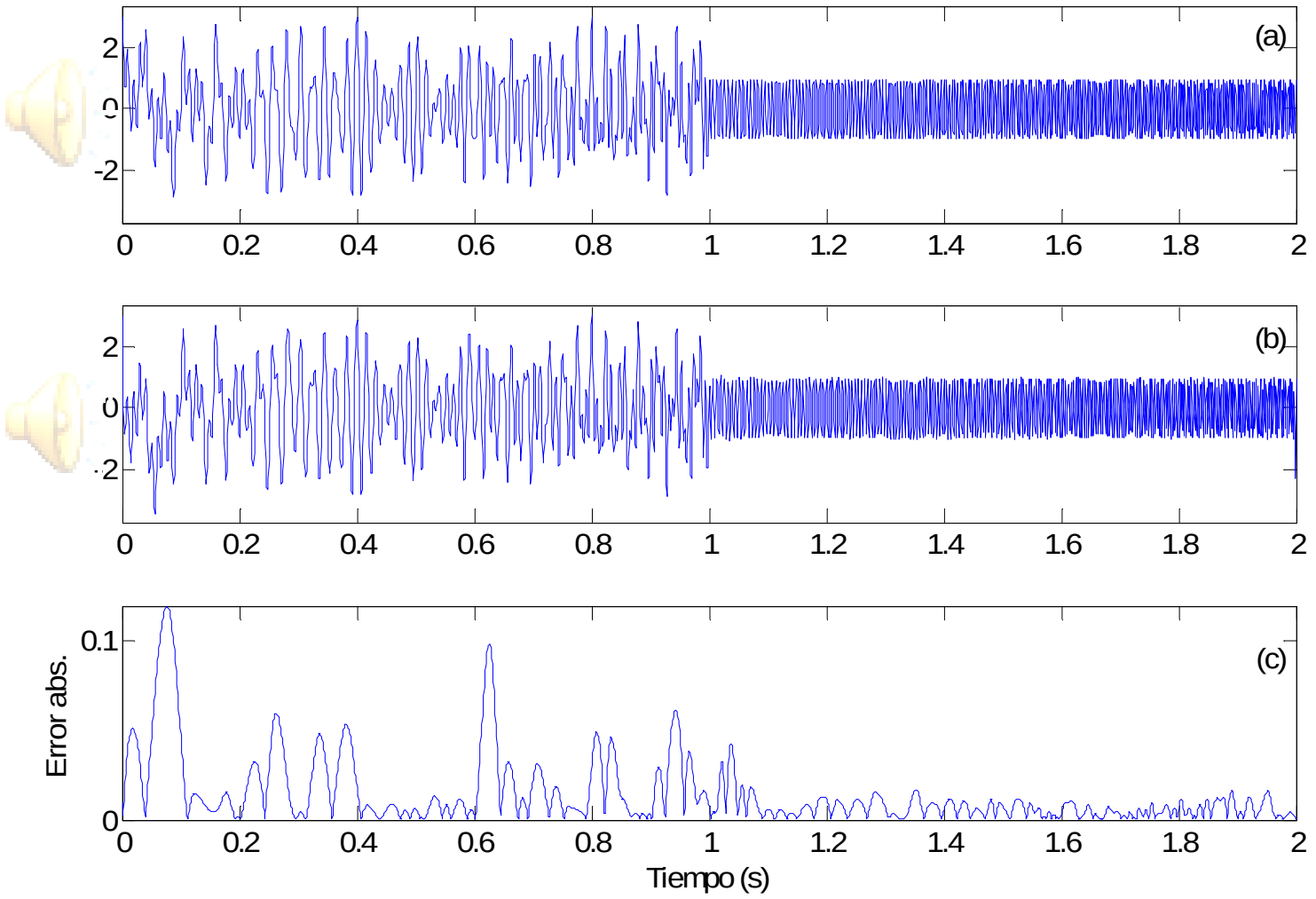
Submuestreo - tasa crítica *local*



Reconstrucción



Entrada vs. reconstruida



¿Compresión eficiente?

- Se determina la cantidad de muestras P_1 y su distribución estadística.
- Para conocer la mínima cantidad de bits/muestra se usa la entropía.

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

- Para saber cuán concentrada está la pdf se utiliza la entropía relativa.

$$H_r = H / H_{max}$$

- # bits = # muestras × H
- Factor de compresión FC = # bits_entrada / # bits_sx_compr.

Resultados de la compresión

$x(t)$ → 2001 muestras x 8 bits/muestra = 16008 bits

	$x(t)$	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	-
# muestras	2001	639	335	170	100	1244
# bits	14267	3431	1578	571	246	5826
FC	1,12	-	-	-	-	2,75

Ahorra 64 % de bits

Más resultados

	Extremos locales			
	|IMF1|	|IMF2|	|IMF3|	Total
H (bits/muestra)	4,49	4,06	2,91	-
H_r	0,48	0,48	0,39	-
# bits	2869	1360	495	4970
FC	-	-	-	3,22

Comprime 15 % más

Conclusiones

- Se introdujo un nuevo método de codificación de señales de audio sin pérdidas basado en EMD y submuestreo adaptativo.
- Se evaluó su rendimiento para una señal particular mediante el factor de compresión alcanzado con codificación por entropía sin memoria.
- Se puede compactar considerablemente la señal con este método.
- La compresión mejora aún más si se evalúan los valores absolutos de los extremos.

Trabajo futuro

- Reducción del ancho de banda de cada IMF para reducir la cantidad de extremos resultantes.
- Diseño de funciones de transformación a los histogramas para suprimir amplitudes no existentes de los extremos.
- Minimización del error de reconstrucción.
- Extensión del método a señales temporales de longitud arbitraria. Análisis en cuadros de longitud variable*.

* Roveri. “Estudio y comparación de codificadores de audio sin pérdidas”, Proyecto Final de Ing. Electrónica, UNR, 2011.


¡Gracias!