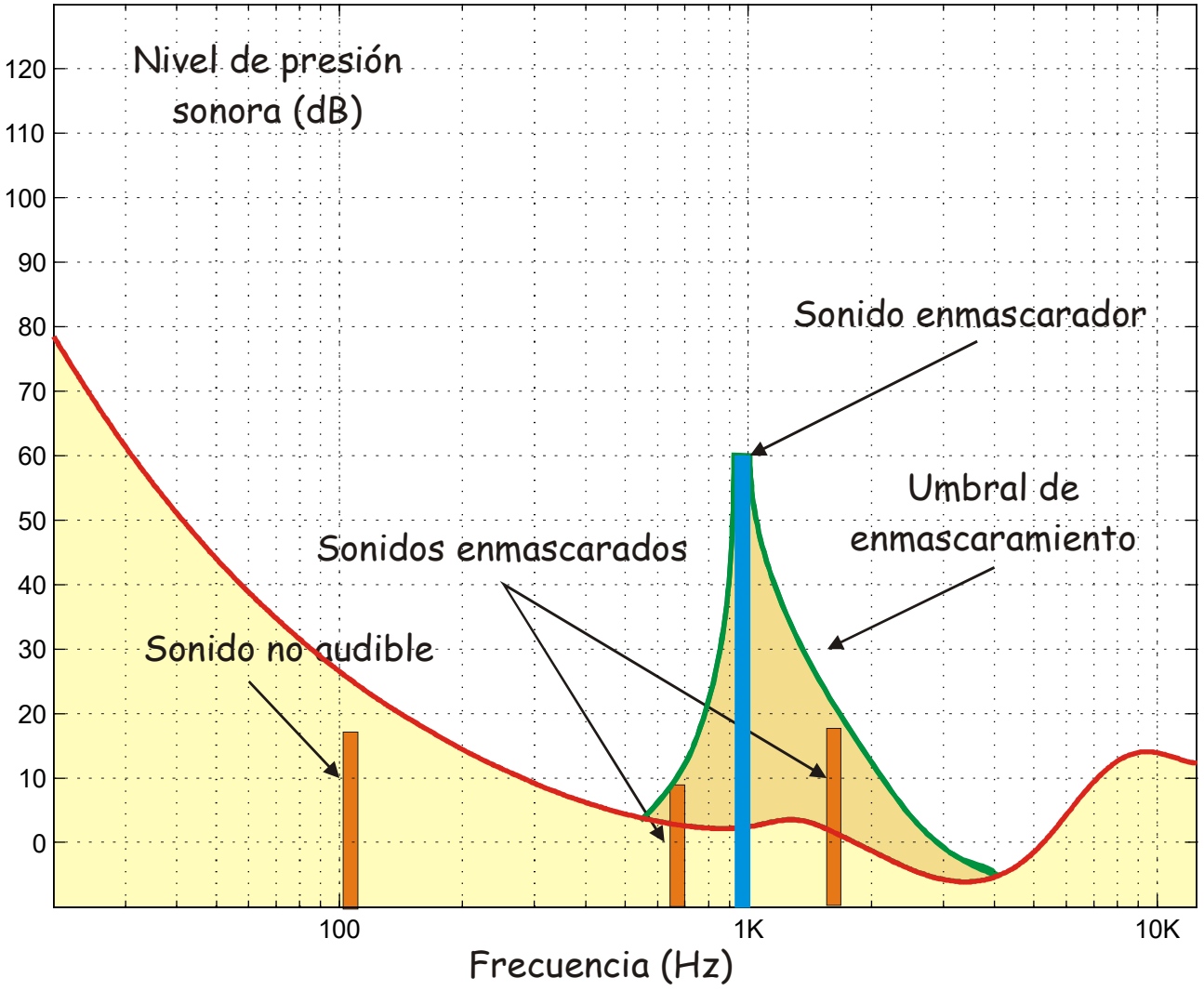
TV: TeleVisión – Plan 2010

Audio en normas MPEG

Enmascaramiento

- El oído interno realiza un análisis “espectral” de corto plazo de manera que diferentes regiones de la membrana basilar se excitan dependiendo de la frecuencia del sonido.
- El sistema auditivo puede ser burdamente descrito como un conjunto de filtros paso banda, fuertemente solapados, con bandas de paso comprendidas entre 50Hz y 100Hz para señales por debajo de 500Hz y hasta 5000Hz para las altas frecuencias
- Deben considerarse hasta 26 bandas críticas, cubriendo un ancho de banda de 24Khz.
- Una señal es inaudible si su nivel es inferior al umbral de audibilidad, que depende de la frecuencia
- El enmascaramiento simultaneo es un fenómeno por el cual una señal de menor nivel (la enmascarada) puede volverse inaudible (ser enmascarada) por la ocurrencia simultanea de otra señal más fuerte (enmascaradora) siempre y cuando ambas señales (enmascarada y enmascaradora) estén suficientemente cercanas en frecuencia.
- Puede medirse un “umbral de enmascaramiento” , de manera que señales por debajo de este nivel no serán audibles.

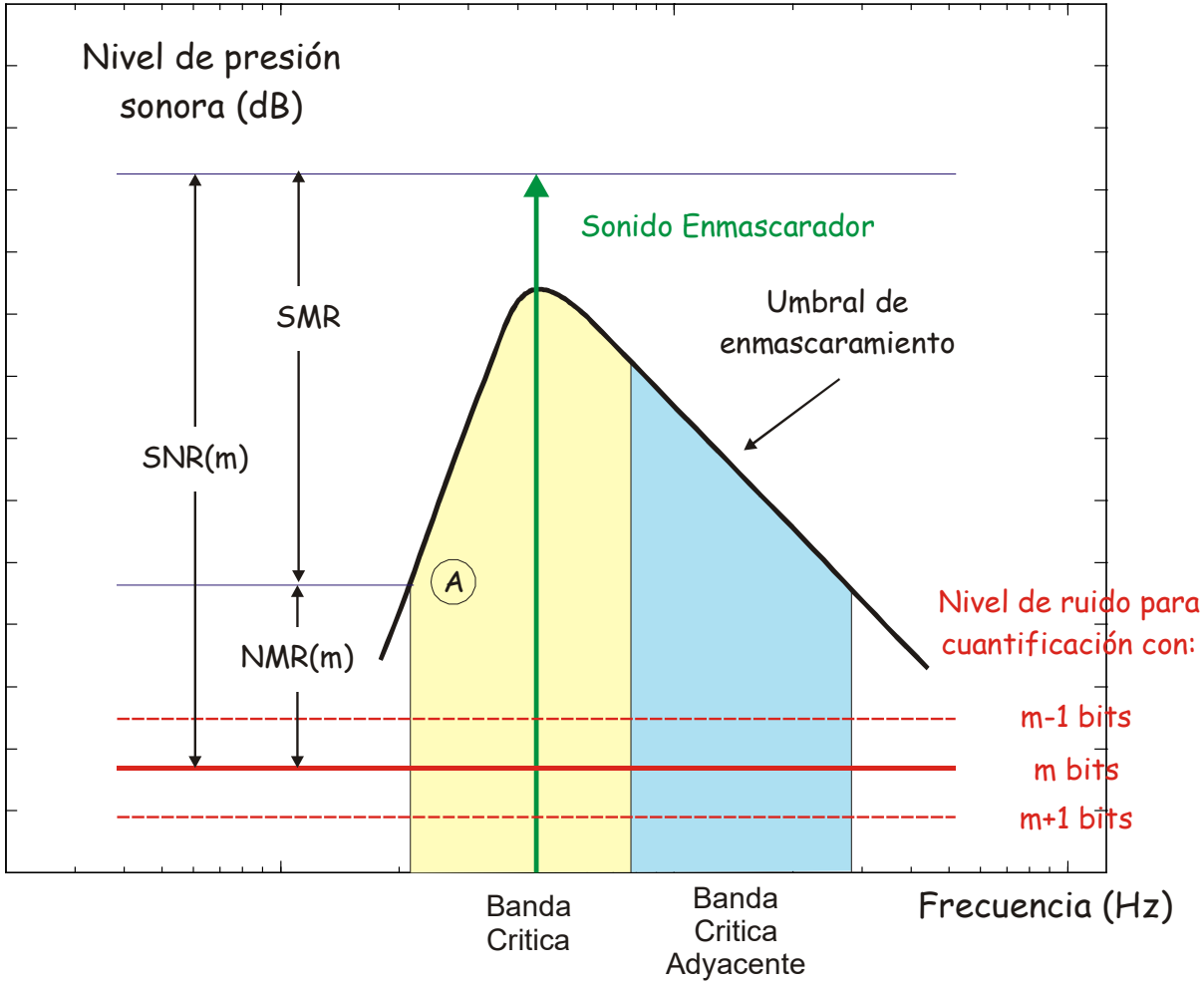
Enmascaramiento



Enmascaramiento simultáneo

- El mayor enmascaramiento se produce en la banda crítica en la que se sitúa el sonido enmascarador, y es menor en las bandas críticas adyacentes
- Las señales cuyo nivel se sitúa por debajo del umbral de enmascaramiento no se oyen
- La señal enmascarada puede ser ruido de cuantificación, distorsión de solapamiento, componentes de la señal de bajo nivel, etc…
- El umbral de enmascaramiento (también conocido como JND -"just noticeable distorsion"-) varía con el tiempo. Depende del nivel de presión sonora del enmascarador, su frecuencia, las características de las señales enmascaradoras y enmascaradas (tono puro, banda estrecha…), etc…
- La pendiente del umbral de enmascaramiento es más abrupta hacia las bajas frecuencias que hacia las altas (esto es, las altas frecuencias se enmascaran más fácilmente)
- La distancia entre entre enmascarador y enmascarado es más pequeña cuando un "ruido" enmascara a un tono que al revés, esto es, el ruido es mejor enmascarador que un tono.

Enmascaramiento del ruido de cuantificación



Enmascaramiento del ruido de cuantificación

- La distancia entre el nivel de la señal enmascaradora y el umbral de enmascaramiento se conoce como “relación señal a máscara” (SMR). Su máximo valor se encuentra en el extremo izquierdo de la banda crítica y su mínimo alrededor de la frecuencia del sonido enmascarador.

- Siempre que la relación señal a ruido de cuantificación (SNR) sea mayor que la SMR el ruido de cuantificación no será audible. Obviamente la SNR depende del número de bits empleados en la cuantificación por lo que denominamos SNR(m) a la SNR resultante de cuantificar con m bits.

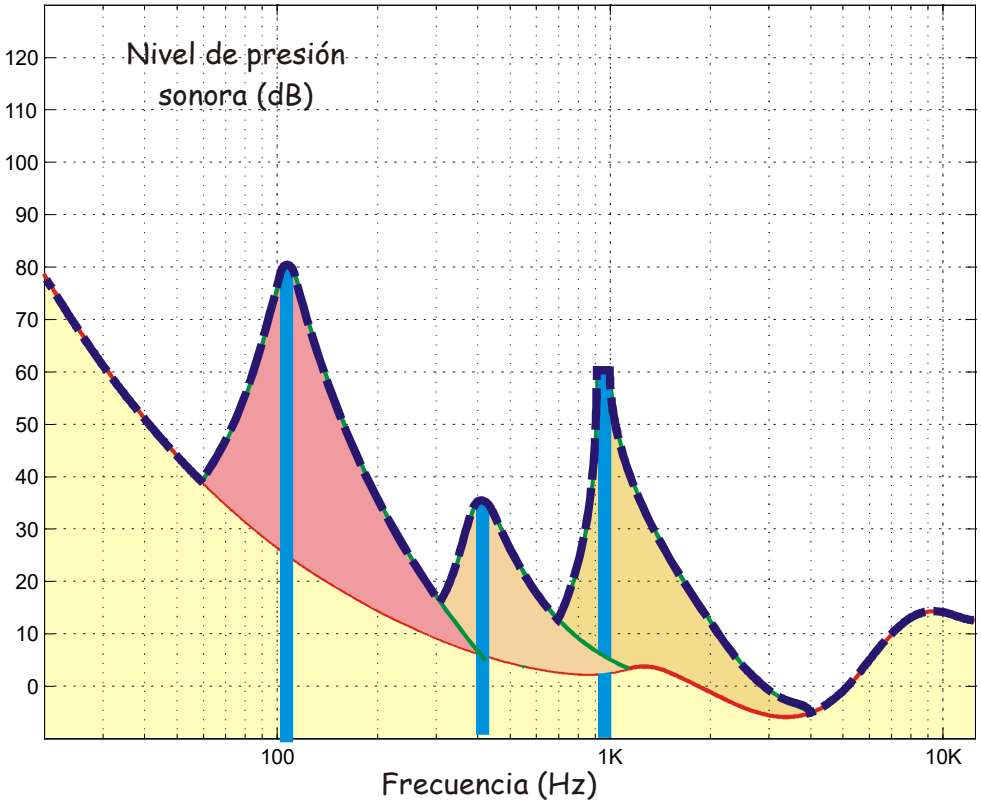
- La distorsión perceptible en una determinada banda se mide por la relación ruido a máscara (NMR):

$$NMR(m) = SMR - SNR(m)\ dB$$

- La NMR representa la distancia entre el nivel de ruido de cuantificación y el nivel que sería audible en una sub-banda determinada. Dentro de una misma banda crítica el ruido de cuantificación no sería perceptible siempre que la NMR(m) sea negativa.

Enmascaramiento simultáneo

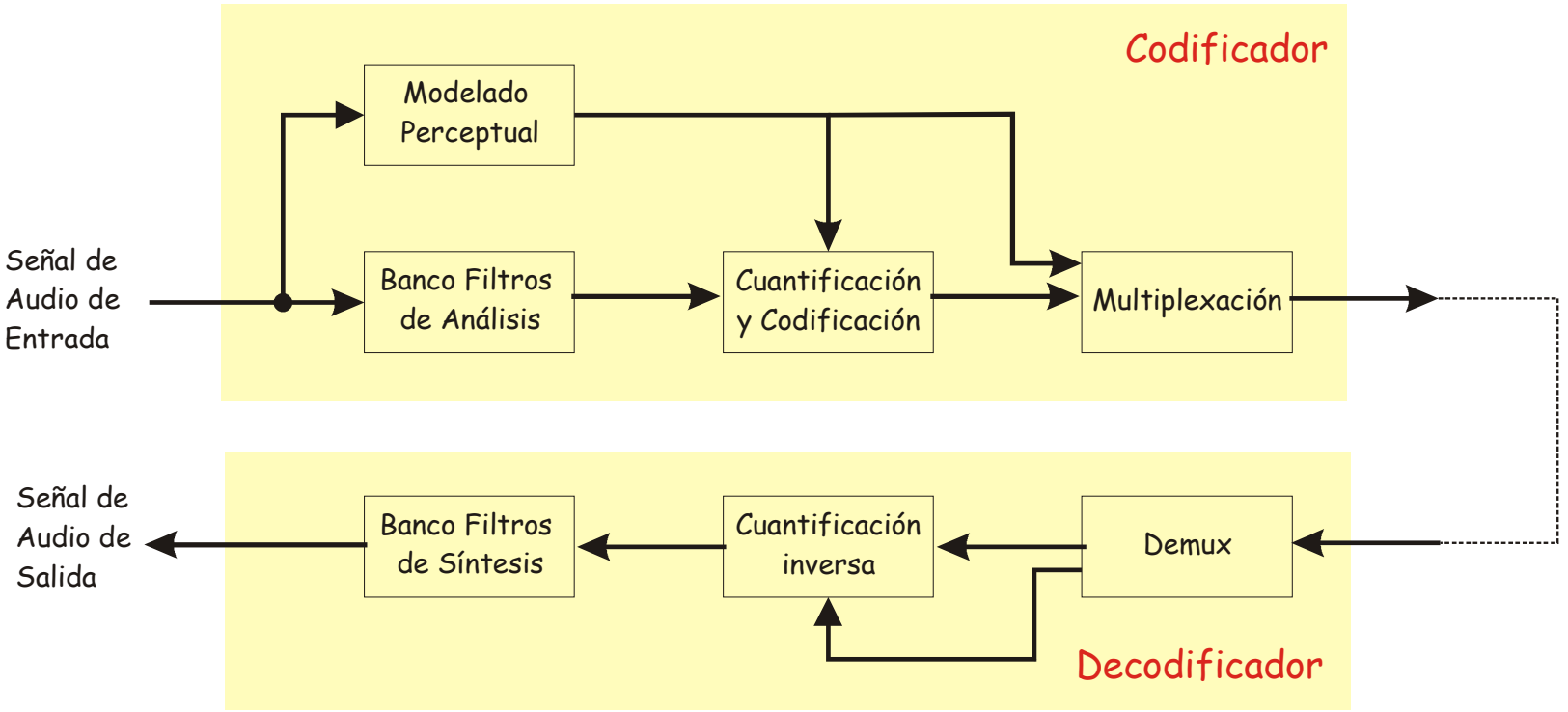
★ Cuando la señal contiene múltiples componentes enmascaradoras, cada una tiene asociado su umbral de enmascaramiento y puede calcularse un umbral de enmascaramiento global que describe el umbral de la mínima distorsión perceptible como una función de la frecuencia



Codificación perceptual

- ★ Utiliza el conocimiento acerca de cómo se percibe la señal (de audio)
- ★ El objetivo no es obtener una buena aproximación a la forma de onda de la señal , sino obtener una señal que sea útil para el receptor humano, para lo cual
  - ✸ Elimina en la medida de lo posible la redundancia existente en la señal, explotando la correlación existente entre las muestras y
  - ✸ Elimina componentes imperceptibles
  - ✸ Explota la diferente respuesta de los sentidos en el dominio de la frecuencia, concentrando el ruido en aquellas bandas de frecuencia donde el ruido es imperceptible o más tolerable (“Noise-shaping”)
- ★ En los codificadores de audio MPEG el perfil de ruido se ajusta dinámicamente para explotar al máximo el enmascaramiento simultaneo y temporal.
- ★ Un codificador perceptual responde básicamente a la siguiente estructura:

Codificación perceptual



Codificación perceptual

- La señal se analiza para obtener una estimación de la SMR con la frecuencia, obteniéndose una estimación de la precisión (bits) del cuantificador necesarios para cada banda. Para ello se realiza en ocasiones un análisis espectral del bloque a codificar. Si la resolución espectral del codificador es suficientemente alta, la estimación del perfil de SMR puede derivarse directamente de los coeficientes de la transformada o de las muestras de cada sub-banda.

- Si la tasa binaria a emplear permite el enmascaramiento total del ruido, el proceso efectuado será totalmente transparente.

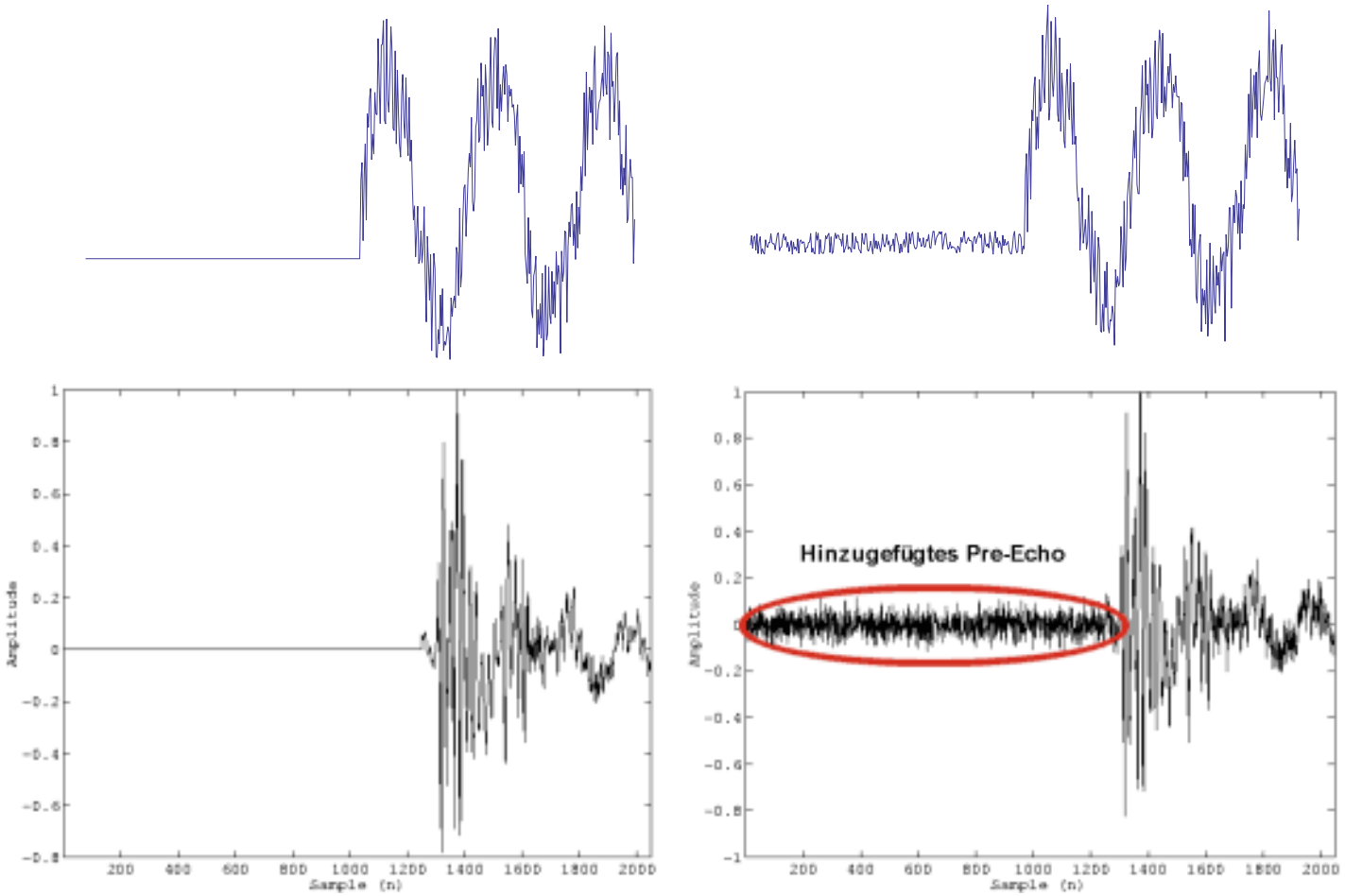
- En la práctica, el post-proceso de la señal por el usuario final, las múltiples etapas de codificación y decodificación que puede sufrir y las propias imperfecciones del modelo de enmascaramiento empleado exigen un margen de seguridad suficientemente amplio.

- El estándar MPEG no es normativo en lo referente al modelado psico-acústico en el codificador, por lo que está abierto a la inclusión de mejores modeladores en el codificador.

Window switching

- ★ Una característica clave de la codificación en el dominio de la frecuencia es la aparición de pre-ecos.
- ★ Considérese un periodo de silencio es seguido por percusión (como castañuelas o triángulos) dentro del mismo bloque.
- ★ El codificador asignará un error de cuantificación significativo y el proceso de reconstrucción diseminará ese error por todo el bloque, originando un pre-eco audible, sobre todo a bajas tasas binarias. Si el ruido no se ha extendido mucha distancia en el tiempo, el pre-eco puede ser enmascarado (pre-enmascaramiento) .
- ★ El pre-eco puede minimizarse o eliminarse haciendo que el tamaño del bloque sea más pequeño. El inconveniente es que la información lateral aumenta drásticamente al reducir el tamaño del bloque.
- ★ Una solución es cambiar dinámicamente el tamaño del bloque, usando bloques pequeños para controlar el pre-eco cuando la señal no es estacionaria

Pre eco (ejemplo)

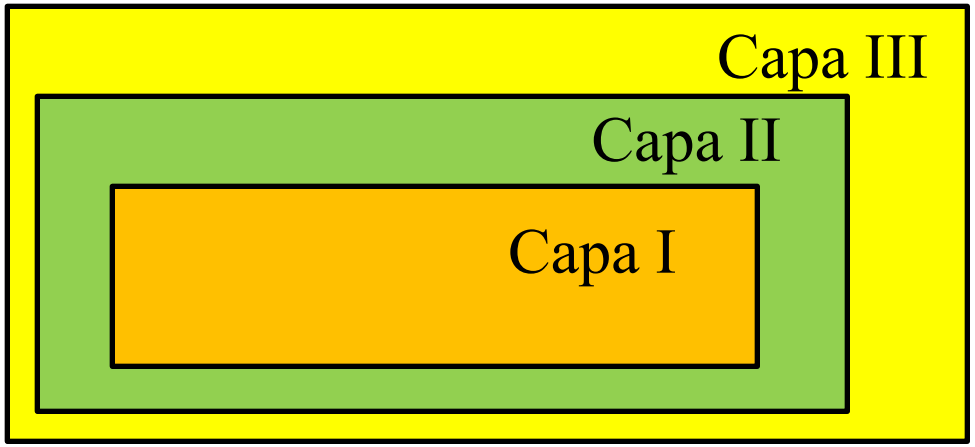


Codificación en el dominio de la frecuencia

- Se utilizan la codificación de transformadas (TC) , la codificación de sub-bandas (SBC) y la codificación hibrida.
  - En la ***codificación de sub-bandas***, el proceso de cuantificación y los propios compromisos tomados en el diseño de los filtros provocan que los términos de solapamiento no se cancelen exactamente, dando lugar a una distorsión por solapamiento. En MPEG/Audio esta distorsión es imperceptible si bien reduce el rango dinámico desde 20 hasta 18 bits.
  - En la ***codificación de transformadas*** se utiliza la DCT o versiones modificadas de la misma (MDCT). La MDCT es una transformada solapada en la que los bloques consecutivos se solapan en un 50%. Tienen una mayor G_{TC}, y los efectos de borde entre bloques son atenuados debido al solapamiento.
  - En los ***codificadores híbridos*** se combinan bancos de filtros y transformadas discretas. Por ejemplo, puede usarse un banco de filtros seguido de una transformada MDCT sobre la salida de cada sub-banda, para obtener una muy alta resolución en frecuencia.
- MPEG-1 utiliza un codificador de sub-bandas para las capas I y II y un codificador hibrido para la capa III.

MPEG-1 Audio

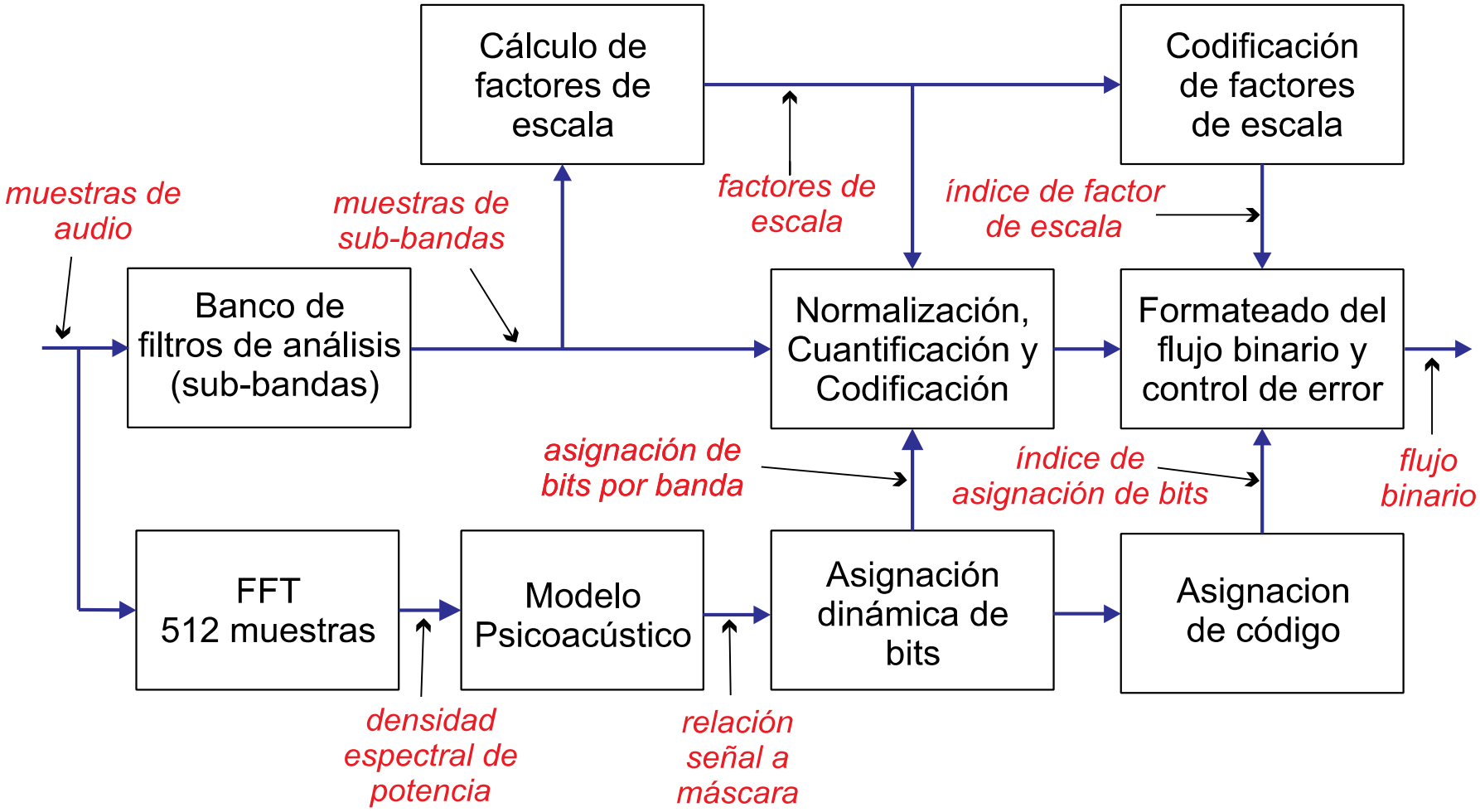
- ★ Codificación de señal de audio monofónica o estéreo con un ancho de banda de hasta 24KHz
- ★ Estándar internacional desde 1993 basado en en dos algoritmos de codificación perceptual desarrollados a finales de los 80.
- ★ Se ha convertido en un estándar universal en múltiples campos (como electrónica de consumo, radiodifusión, telecomunicaciones…)
- ★ Por su amplio rango dinámico puede superar en ocasiones la calidad del CD.
- ★ El estándar fija tres capas (I, II y III) de complejidad, retardo y calidad subjetiva crecientes.
- ★ Las capas superiores incorporan los bloques principales de las capas inferiores.
- ★ Un decodificador MPEG-1 completo es capaz de decodificar las tres capas.



Layer I	**versión simplificada de MUSICAM**
Layer II	**MUSICAM**
Layer III	**Combinación de MUSICAM y ASPEC**

MPEG-1 Audio, Capa I

Codificador Capa I



Banco de filtros de análisis + diezmado

- Sigue el esquema de un codificador perceptual
- En una primera etapa, se cambia la representación de la señal desde el dominio del tiempo al dominio de la frecuencia mediante un banco de 32 filtros de análisis (polifase de 512 coeficientes) de igual ancho de banda.
- Por estar equiespaciados, las bandas de paso de los filtros no coinciden con las bandas críticas.

Para una frecuencia de muestreo de 48KHz:

$$W = f_s / 2 = 24\,\text{KHz} \quad \rightarrow \quad W_i = \frac{24}{32}\text{KHz}=750\ \text{Hz}$$

A bajas frecuencias un mismo filtro cubre varias subbandas

- La capa III aplica una transformación MDCT a la salida de cada subbanda (esquema híbrido)

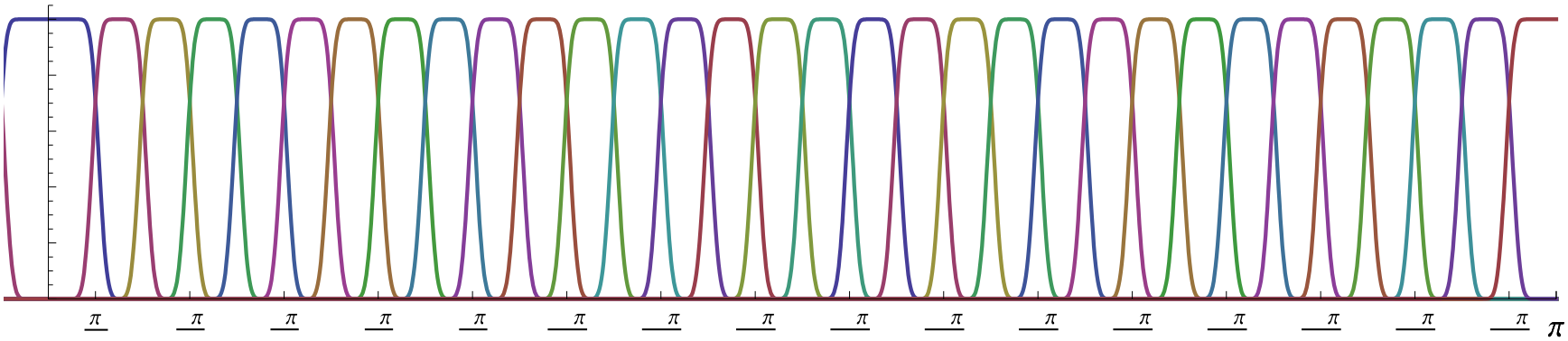
Banco de filtros de análisis + diezmado

- La respuesta al impulso de la sub-banda k, $h_k(n)$, se obtiene modulando un filtro paso bajo prototipo, $h(n)$

$$h_k(n) = h(n) \cdot \cos\left[\frac{2k-1}{2M} + \varphi(k)\right] \quad \text{con } M = 32,\ k = 1,2,\ldots,32,\ n = 1,2,\ldots,512$$

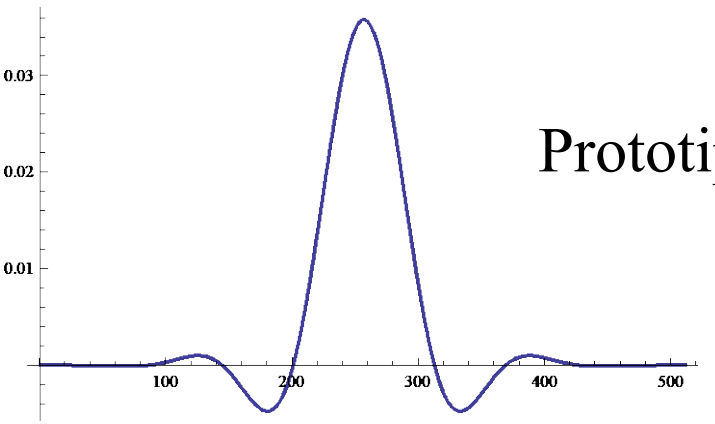
- A 48KHz de frecuencia de muestreo, $h(n)$ tiene un ancho de banda de 3dB de 750/2 Hz=375 Hz, por lo que las salidas de los filtros se solapan, aunque los solapamientos se cancelan entre sí.
- Si no se aceptase ninguna aproximación ni compromiso en los filtros (ver “codificación de subbandas”) y no se realizase ningún tipo de cuantificación sería posible una recuperación básicamente perfecta de la señal. En la práctica se exige a los filtros una caída muy abrupta para evitar el que el ruido de cuantificación afecte a bandas adyacentes (los filtros implementados proporcionan más de 96dB de atenuación de las bandas solapadas)
- La salida de cada filtro es diezmada en un factor 32

Banco de filtros de análisis

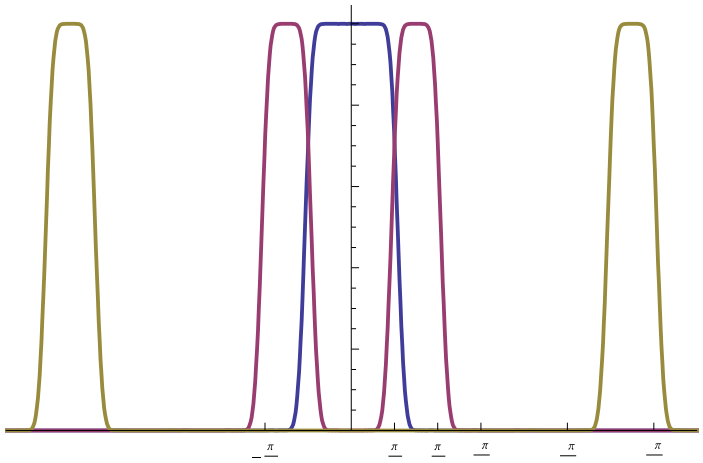


Banco de filtros (frec. 0 a π)

Filtros número 0,1 y 6



Prototipo h(n)



Capa I: escalado

- Para cada sub-banda, después del diezmado, se empaquetan 12 muestras consecutivas en un bloque, dando lugar a 32 bloques en paralelo que almacenan 12x32=384 muestras.

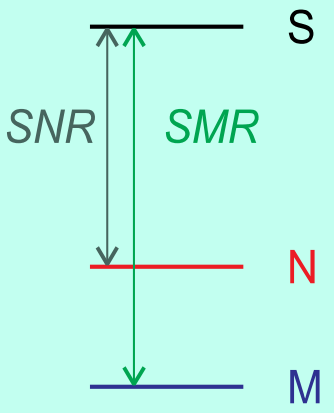
- Para cada bloque (B_n) se escoge un factor de escala (SF_n) de entre 63 posibles de manera que cuando las muestras del bloque se dividen por el factor de escala tengan un valor absoluto inferior, pero lo más próximo posible, a 1.

- Este proceso se realiza usualmente seleccionando en cada bloque la muestra con mayor valor absoluto y buscando a continuación en la tabla de factores de escala el primero inmediatamente superior al valor anterior.

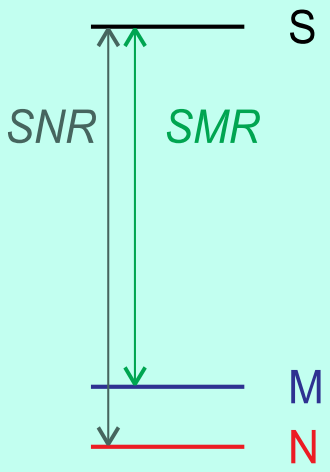
Modelos psicoacústicos

- Los detalles de cómo implementar el modelo psicoacústico no son parte de la norma, pero incluye un anexo informativo con los detalles de implementación de dos modelos psicoacústicos.
- Se sugiere que
  - El modelo psicoacústico I se use para las capas I y II
  - El modelo psicoacústico II se use para la capa III
- El modelo psicoacústico se utiliza para calcular el umbral de enmascaramiento para una señal de entrada dada.
- El umbral de enmascaramiento permite decidir el número de bits necesarios para representar (cuantificar) las muestras sin que el ruido de cuantificación sea audible.
- El algoritmo determina el mínimo umbral de enmascaramiento (esto es, el umbral de enmascaramiento en el caso peor)
- La relación señal a máscara (SMR) de una sub-banda se calcula a partir del mínimo umbral de enmascaramiento y de la potencia máxima de la señal en cada banda.
- Como las señales de cada sub-banda están normalizadas (factor de escala, "*scalefactor*") existe una relación conocida entre el número de bits utilizado en la cuantificación y la relación señal a ruido para esa sub-banda.
- Para garantizar que el ruido de cuantificación no es audible simplemente se requiere emplear los bits suficientes para que SNR>SMR

Modelos psicoacústicos



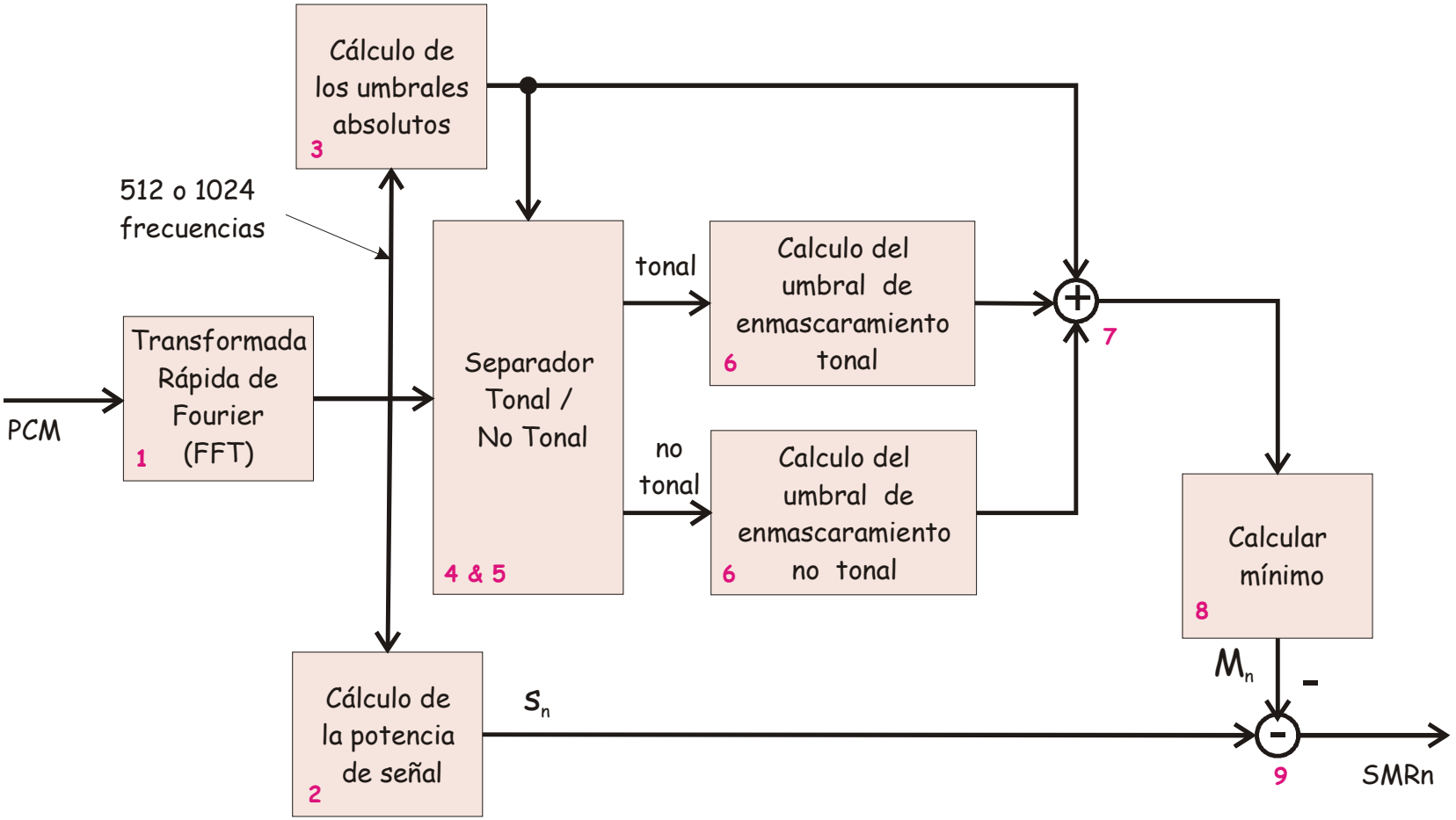
Ruido audible
SNR<SMR



Ruido inaudible
SNR>SMR

Para garantizar que el ruido de cuantificación no es audible simplemente se requiere emplear los bits suficientes para que SNR>SMR

Modelo psicoacústico I



Modelo psicoacústico I

El modelo psicoacústico I determina el valor de SMR en 8 pasos:

1. Cálculo de la FFT de la señal de entrada
2. Determinación del máximo nivel de presión sonora en cada sub-banda
3. Determinación del umbral absoluto de audición
4. Identificación de las componentes tonales (más parecidas a una sinusoide) y no tonales (más parecidas al ruido)
5. Diezmado de los enmascaradores, conservando sólo los más relevantes
6. Cálculo de los niveles de enmascaramiento individuales
7. Determinación del umbral global de enmascaramiento
8. Determinación del umbral mínimo de enmascaramiento para cada sub-banda

A continuación vamos a introducir el modelo usando como ejemplo los resultados obtenidos al procesar las muestras 24961 a 25344 de una grabación del primer movimiento de la Sinfonía de los Juguetes en Do mayor de Leopoldo Mozart.

MP-I: cálculo de la FFT

- La densidad espectral de potencia de la señal se calcula usando una FFT sobre la señal de entrada (de 512 puntos para la capa I y de 1024 para la capa II), "enventanada" con una ventana de Hann:

$$h(i) = \sqrt{\frac{8}{12}}\left[1 - \cos\left(2\pi\frac{i}{N}\right)\right] \quad 0 \le i \le N-1$$

- Para la capa I, debido al retraso introducido por el filtro de análisis, para que la FFT corresponda con el grupo de muestras que se va a codificar debe introducirse un retardo de 256 muestras en la FFT, y para que la ventana de Hann esté correctamente alineada, debe introducirse un retardo adicional de 64 muestras, de forma que el retraso total es de 320 muestras.

- La densidad espectral de potencia es entonces:

$$\log_{10}\left[\left|\frac{1}{N}\sum_{i=0}^{N-1} h(i)\,x(i)\,e^{-\mathrm{j}\frac{2\pi i\cdot k}{N}}\right|\right] \text{ dB}, \quad \text{con } k = 0,1,\ldots\frac{N}{2}-1$$

- Para tener en cuenta las variaciones en el rango dinámico de la señal de entrada la densidad espectral de potencia se normaliza de manera que la máxima amplitud corresponda a un nivel de presión sonora (SPL) de 96 dB.

Máximo nivel de presión sonora por sub-banda

El nivel de presión sonora en cada sub-banda se calcula mediante la expresión:

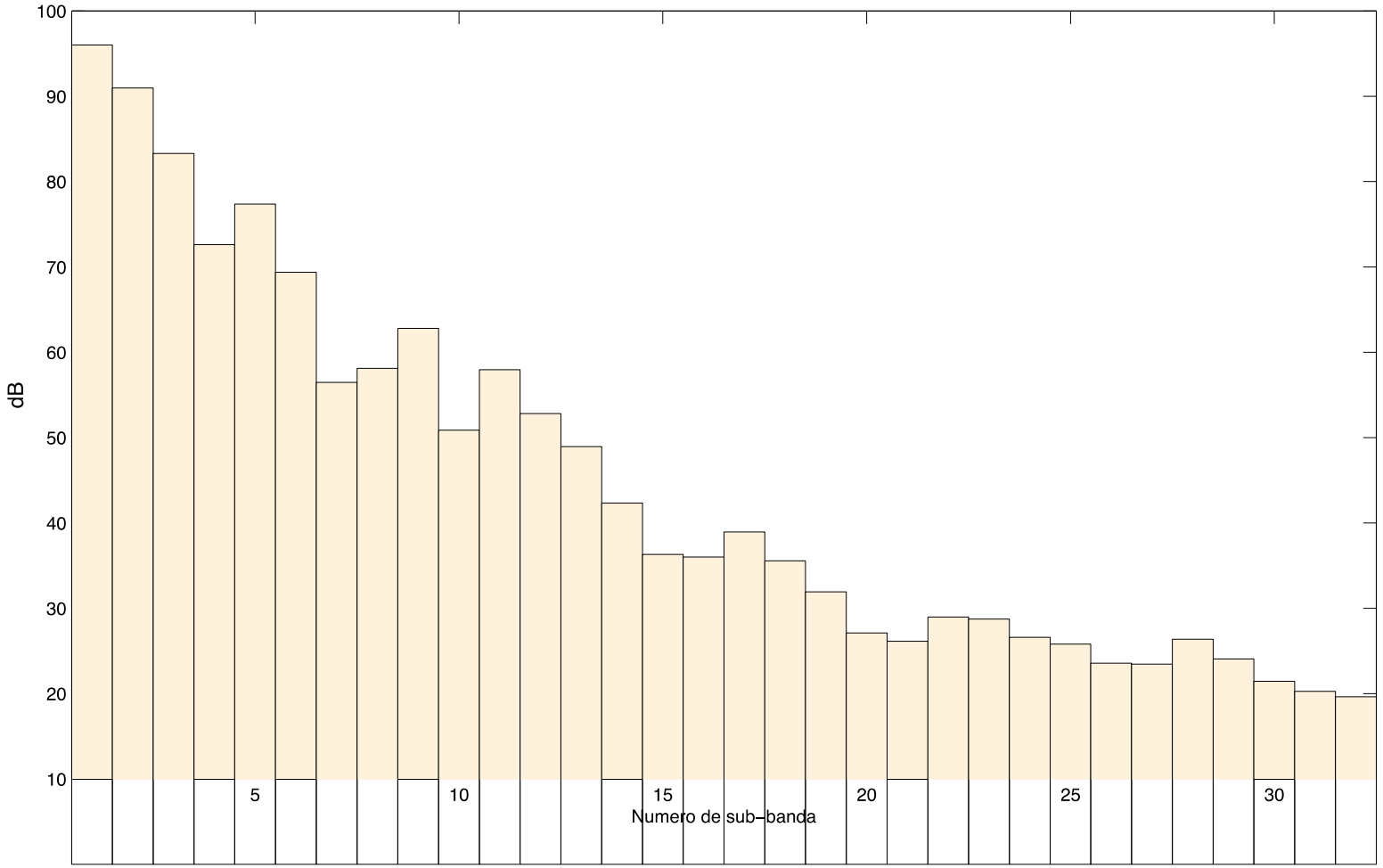
$$\max_{X(k)\text{ en sub-banda n}} \left[X(k), 20\log_{10}\left(32768 \cdot \text{scf}_{\max}(n)\right) - 10 \right]$$

En la capa I scf(n) es simplemente el factor de escala ("scalefactor") de la sub-banda, mientras que en la capa II es el máximo de los tres factores de escala que corresponden a cada uno de los tres grupos de 12 muestras consecutivas.

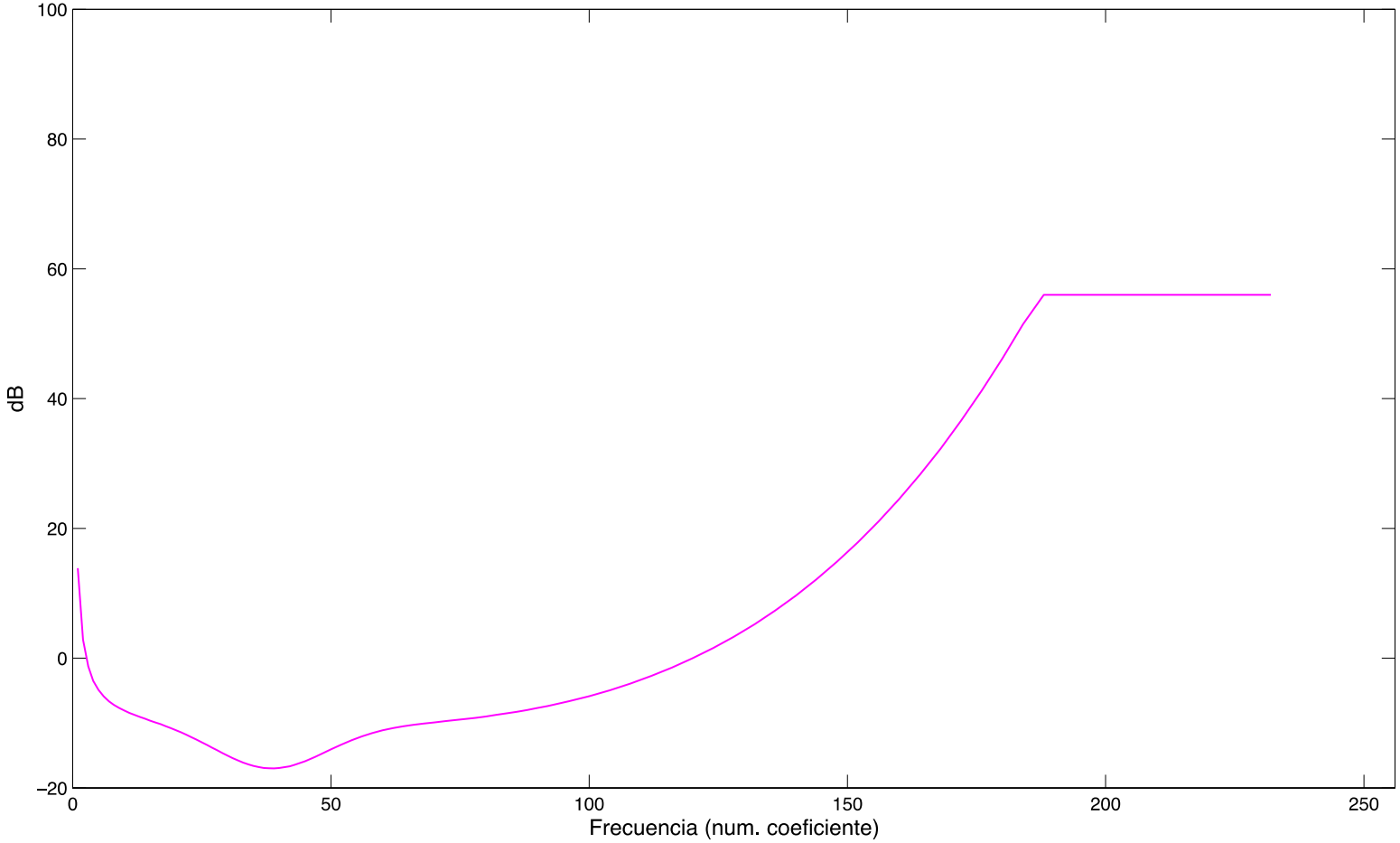
Por ejemplo, para la capa I , el numero de valores X(k) a examinar en cada sub-banda es

$$\frac{256 \text{ valores X(k)}}{32 \text{ sub-bandas}} = 8$$

Máximo nivel de presión sonora por sub-banda



Umbral absoluto de audibilidad



Búsqueda de los enmascaradores tonales

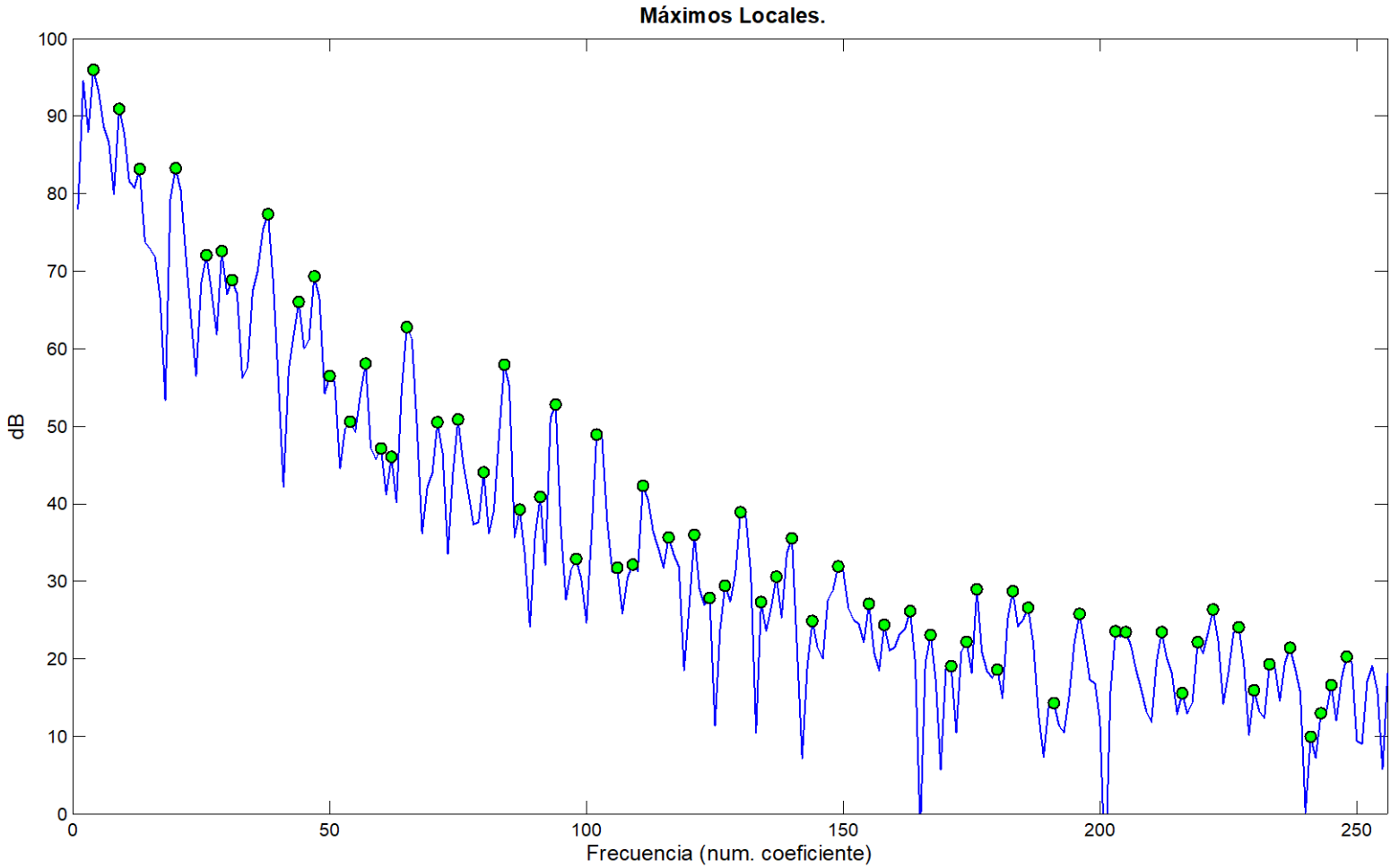
★ Un enmascarador tonal debe corresponder a un pico estrecho en la densidad espectral de potencia de la señal. En consecuencia, el primer paso es encontrar los máximos locales de X(k). Una línea X(k) se considera un máximo si

$$X(k - 1) < X(k) \text{ y } X(k) > X(k + 1)$$

★ Para considerarse un enmascarador tonal el máximo local debe tener una amplitud significativamente mayor que su entorno espectral. Además, el sistema auditivo humano tiene una mejor resolución en frecuencia a frecuencias bajas. En otras palabras, dos tonos que se perciben como frecuencias separadas a bajas frecuencias pueden no serlo a frecuencias más altas. Por esta razón el número de componentes espectrales vecinos que se comparan para decidir si un máximo local es o no un enmascarador tonal aumenta con la frecuencia.

★ En la siguiente figura se muestran los máximos locales obtenidos en el fragmento de la señal ejemplo.

Búsqueda de los enmascaradores tonales: Máximos locales



Identificación de enmascaradores tonales

Un máximo local es considerado un enmascarador tonal si

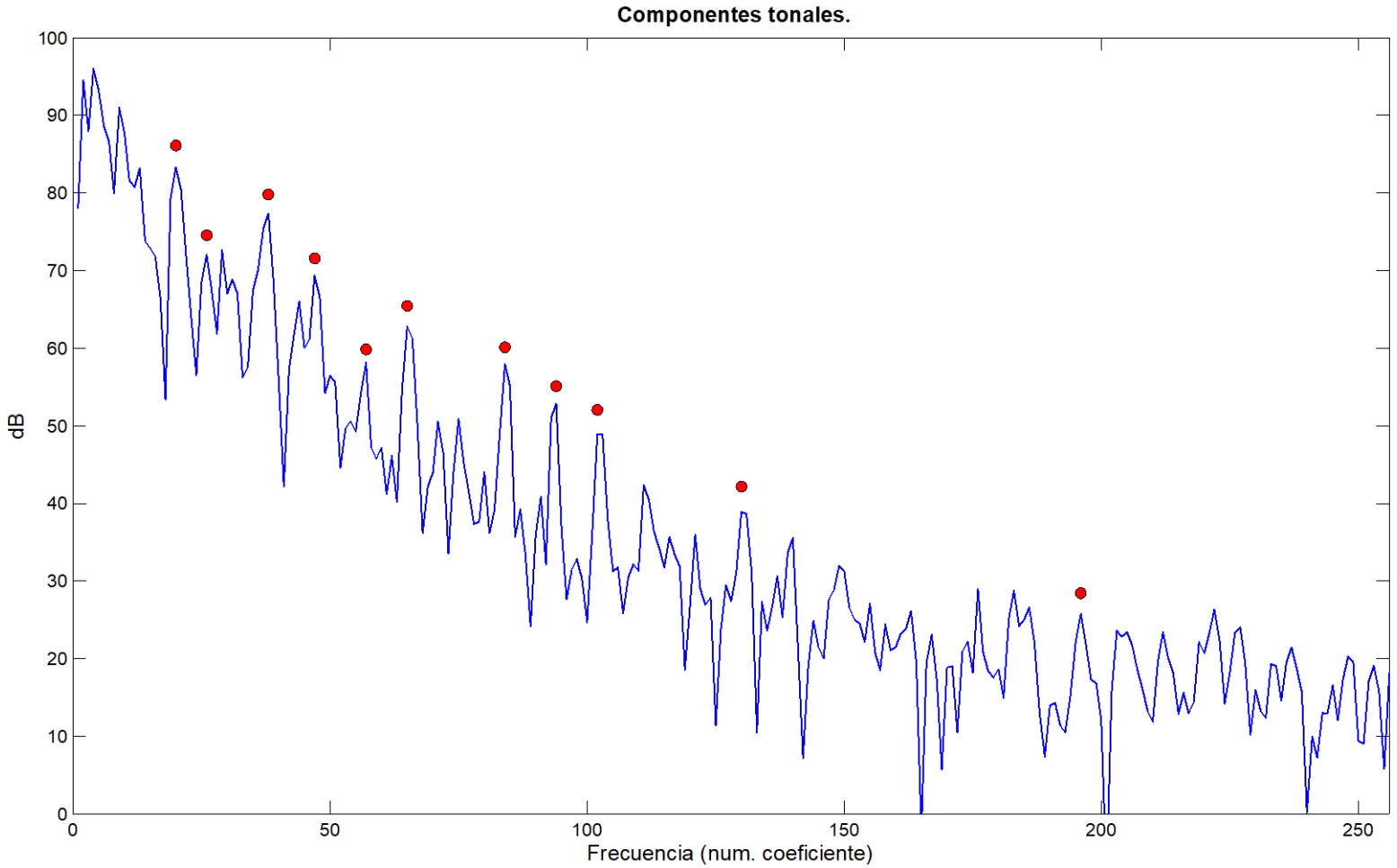
$$X(k) - X(k-j) \geq 7\text{dB donde:}$$

Capa I		Capa 2	
j	rango	j	rango
-2,2	$2 < k < 63$	-2,2	$2 < k < 63$
-3,-2,2,3	$64 \leq k < 127$	-3,-2,2,3	$64 \leq k < 127$
$-6,\ldots,-2,2,\ldots 6$	$128 \leq k < 250$	$-6,\ldots,-2,2,\ldots 6$	$128 \leq k < 255$
		$-12,\ldots,-2,2,\ldots 12$	$256 \leq k < 500$

Si un máximo local satisface las condiciones anteriores, el nivel de presión sonora asignado es:

$$X_{tm}(k) = 10\log_{10}\left(10^{\frac{X(k-1)}{10}} + 10^{\frac{X(k)}{10}} + 10^{\frac{X(k+1)}{10}}\right) \text{ dB}$$

Identificación de enmascaradores tonales



Enmascaradores no tonales

- En primer lugar, los enmascaradores tonales no vuelven a considerarse al estimar el nivel sonoro de los enmascaradores no tonales, por lo que las componentes espectrales en la vecindad de un enmascarador tonal se suprimen (nivel= -∞).

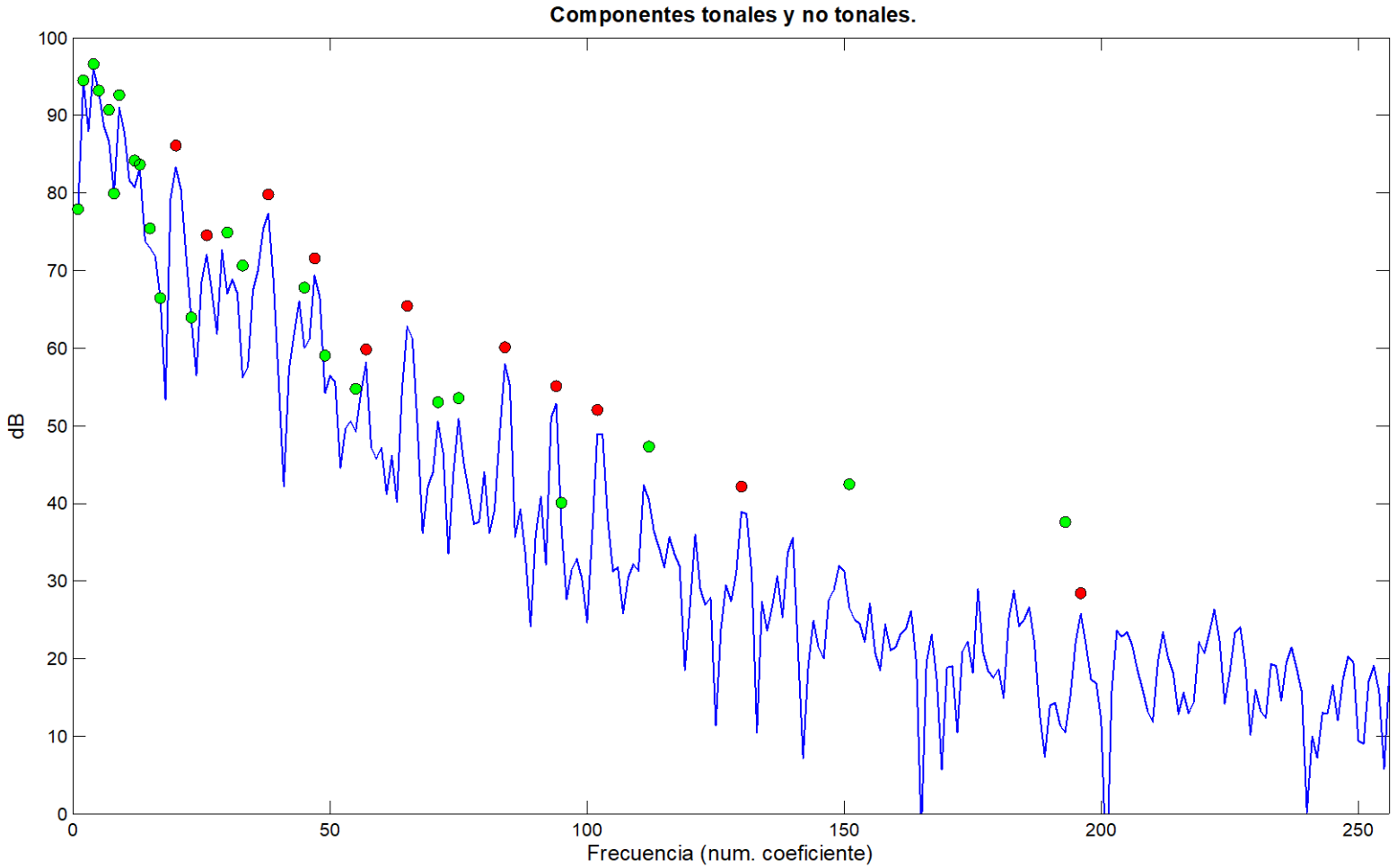
- El nivel de presión sonora de los enmascaradores no tonales se calcula a partir de las líneas espectrales que quedan después de haber suprimido las líneas espectrales asociadas al enmascarador tonal.

- Para ello se divide el rango de frecuencia en bandas críticas y para cada banda crítica se calcula la suma de la presión sonora de cada componente espectral después de haber eliminado los componentes tonales:

$$X_{nm}\left(k_{cb}\right) = 10\log_{10}\left(\sum_{k=k_{lo}}^{k_{up}} 10^{\frac{X(k)}{10}}\right) \text{ dB}$$

donde k_{lo} representa el índice del primer coeficiente dentro de la banda crítica en cuentión y k_{up} el último. La posición del componente notonal dentro de la banda crítica, k_{cb}, se asigna al índice que está más cerca de la media geométrica de los índices de la banda crítica.

Enmascaradores no tonales



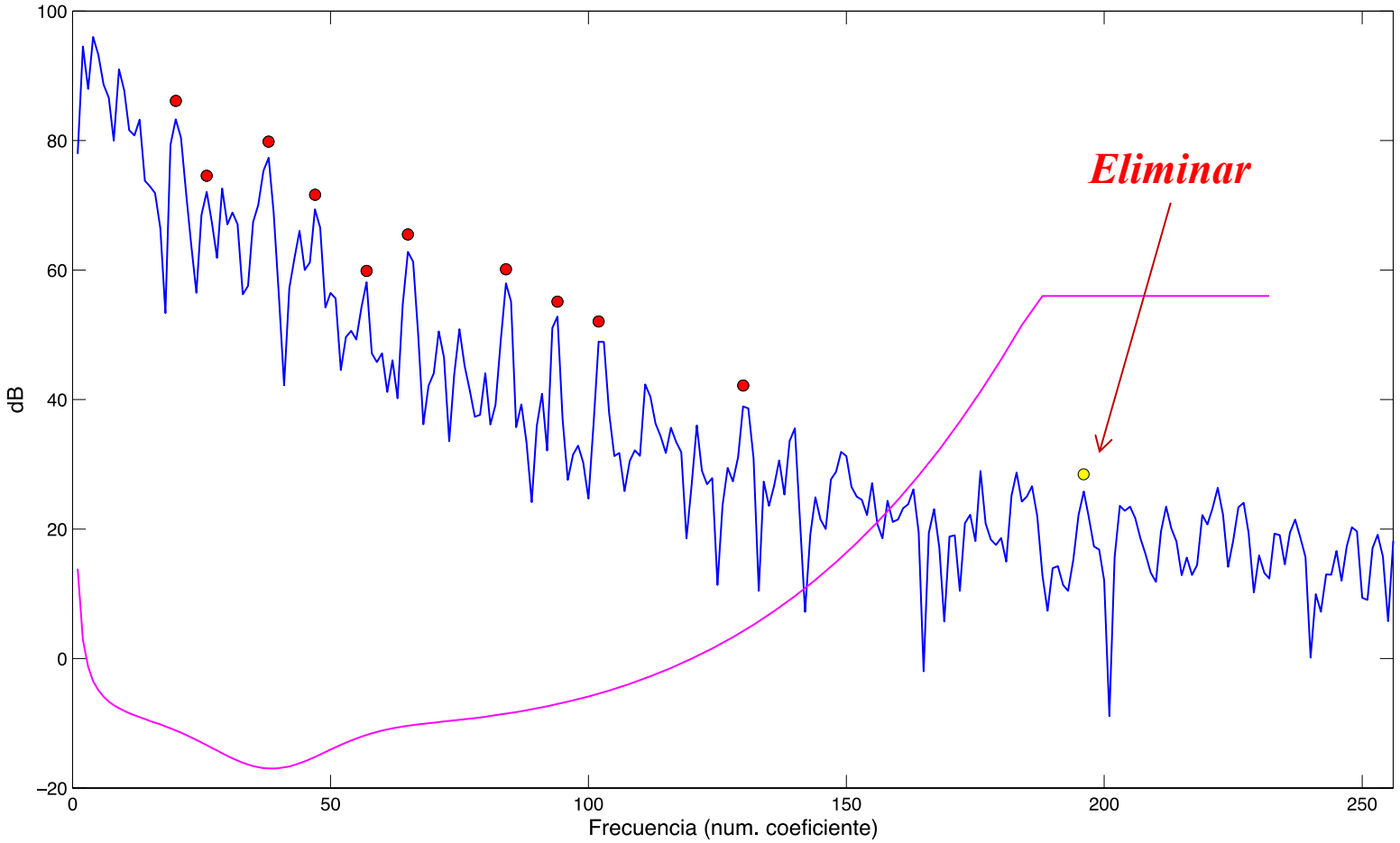
Diezmado de componentes tonales y no tonales

★ Los componentes tonales y no tonales que están por debajo del umbral absoluto son eliminados, esto es, solo se conservan los enmascaradores tonales y no tonales que verifican

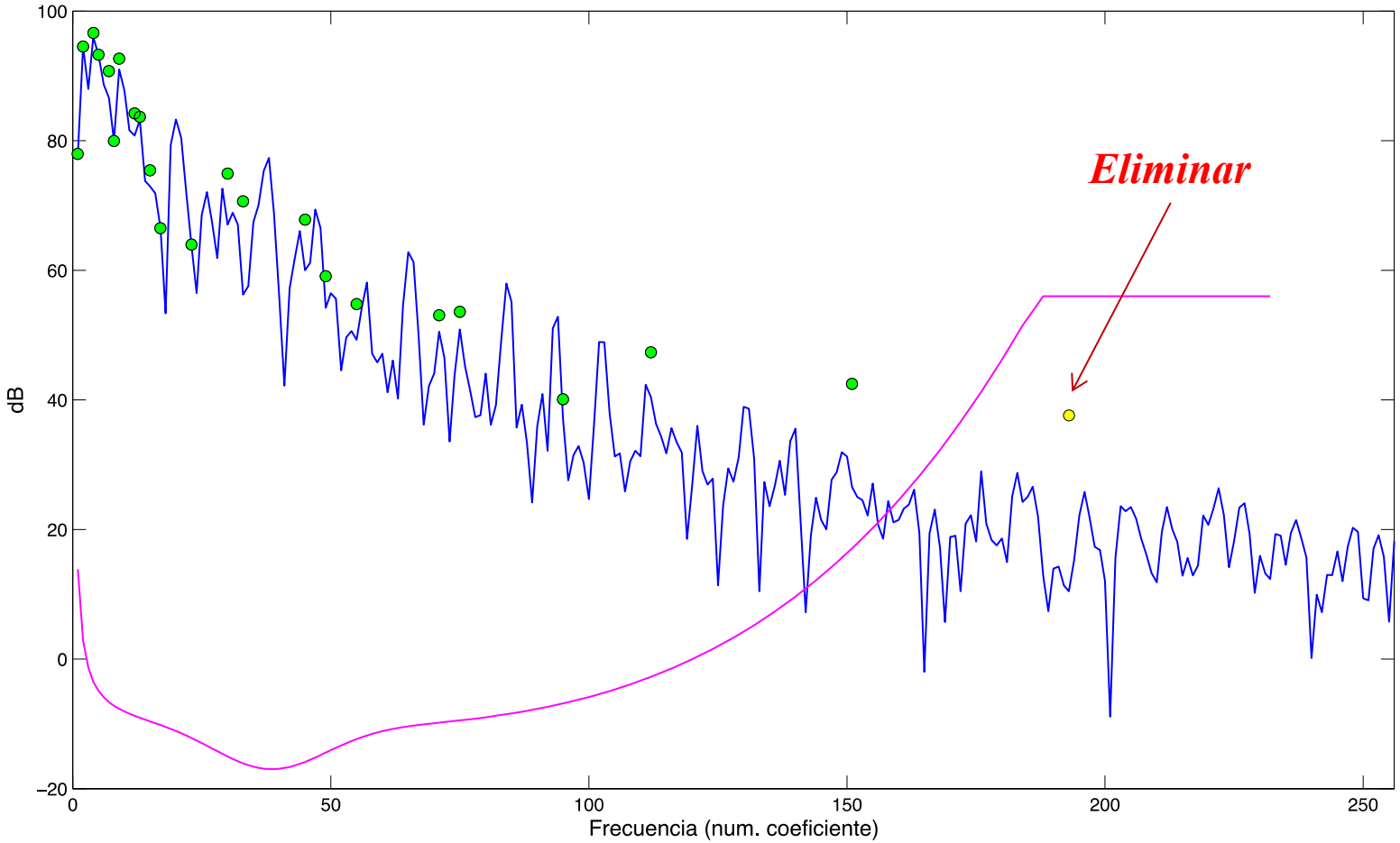
$$X_{tm}(k) \geq LT_q(k) \quad \text{y} \quad X_{nm}(k) \geq LT_q(k)$$

★ Además, si dos enmascaradores tonales están muy próximos (a una distancia menor de 0.5 Bark) se elimina el enmascarador de menor potencia.

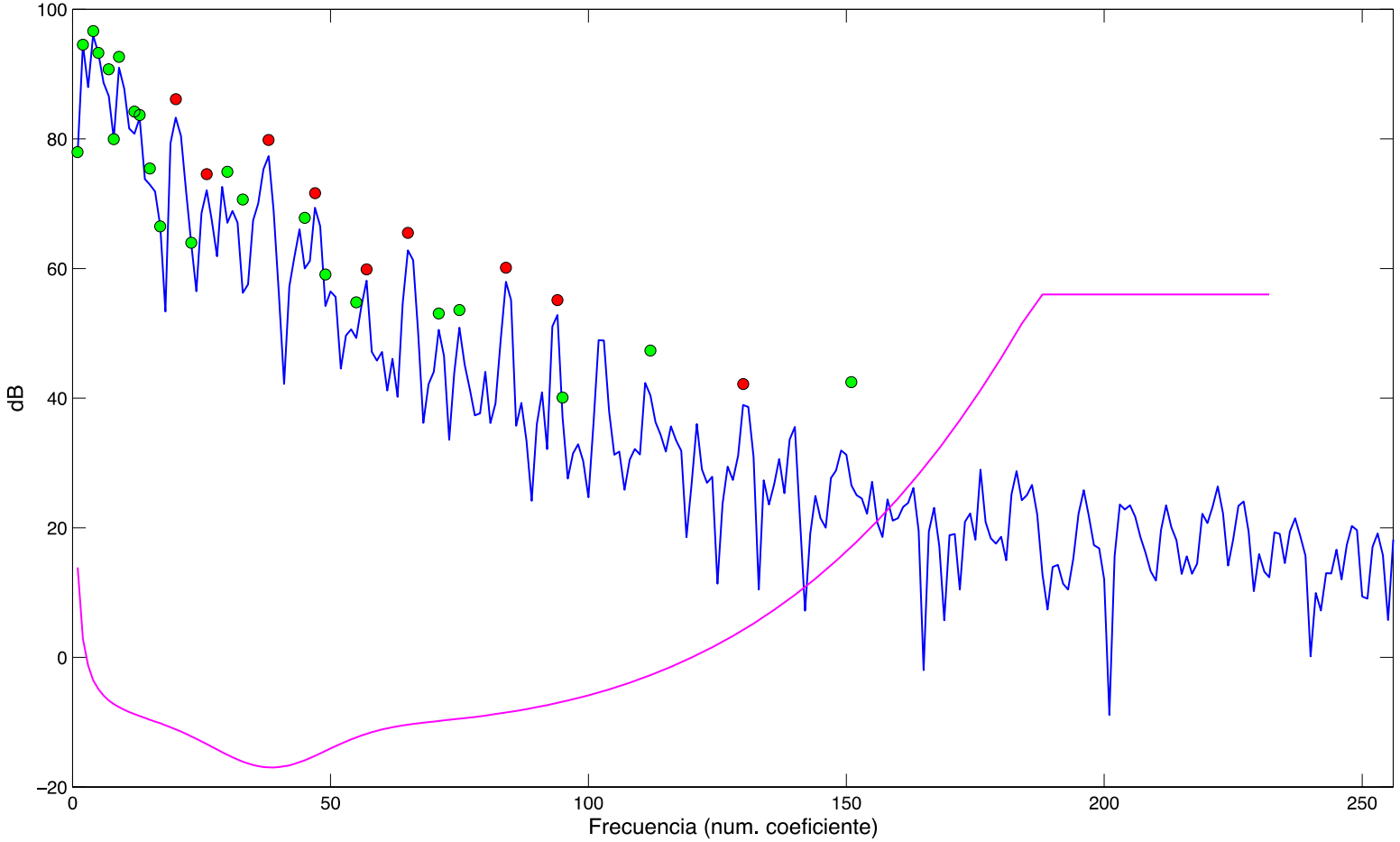
Diezmado de enmascaradores tonales



Diezmado de enmascaradores no tonales



Diezmado de enmascaradores (resultado)



Umbrales de enmascaramiento individuales

El umbral de enmascaramiento individual para el enmascarador tonal situado en la frecuencia con índice j está dado por:

$$LT_m(i,j) = X_{tm}(k) + av_{tm}(k) + vf(k,j)$$

donde k es el índice en frecuencia del i-esimo enmascarador considerado, $av_{tm}(k)$ es el índice de enmascaramiento para enmascaradores tonales:

$$av_{tm}(k) = -0.275 \cdot z(k) - 6.025$$

donde z(k) representa la posición del enmascarador medido en Barks:

$$z(k) = \frac{28 f(k)}{f(k) + 2200} - 0.5 \quad \text{Bark}$$

donde, a su vez, f(k) es la frecuencia correspondiente al índice k. El índice de enmascaramiento establece un nivel constante en el umbral de enmascaramiento para un enmascarador.

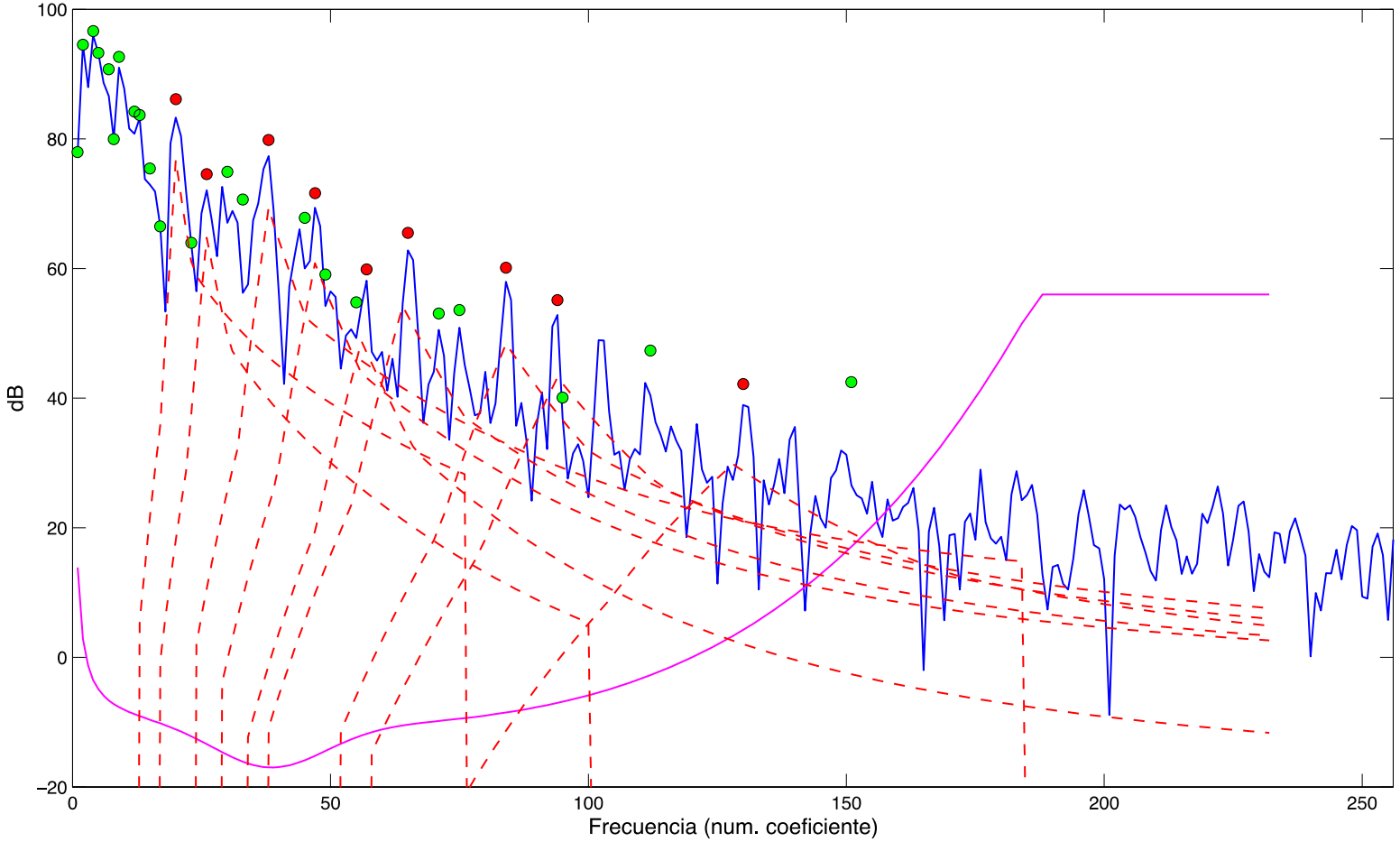
Umbrales de enmascaramiento individuales

La función de enmascaramiento define la pendiente del umbral de enmascaramiento para las frecuencias por encima o por debajo del enmascarador:

$$vf(k,j)=\begin{cases} 17(dz+1)-(0.4X(k)+6) & \text{si } -3\le dz<-1 \\ (0.4X(k)+6)dz & \text{si } -1\le dz<0 \\ -17dz & \text{si } 0\le dz<1 \\ -(dz-1)(17-0.15X(k))-17 & \text{si } 1\le dz<8 \end{cases}$$

donde dz es la distancia en Bark del enmascarador: $dz=z(j)-z(k)$. Los umbrales de enmascaramiento para índices que se encuentran situados a distancias inferiores a -3 Bark o superiores a 8 Bark se sitúan en $-\infty$

Umbrales de enmascaramiento individuales



Umbrales de enmascaramiento indiv. no tonales

Para enmascaradores no tonales la expresión del umbral de enmascaramiento es similar:

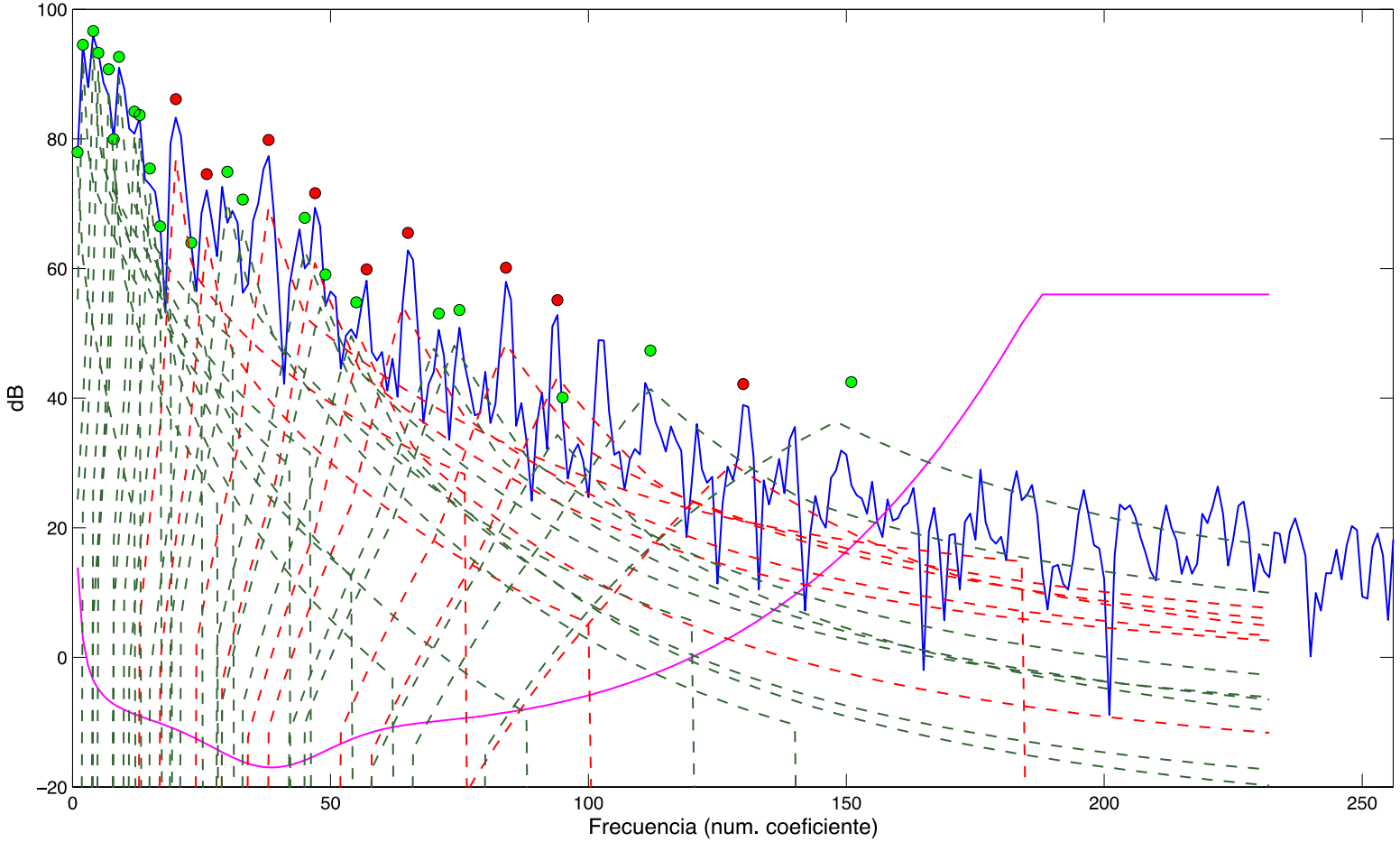
$$LT_{nm}(i,j) = X_{nm}(k) + av_{nm}(k) + vf(k,j)$$

donde k es el índice de frecuencia (número de coeficiente) del iésimo enmascarador considerado, y $av_{nm}(k)$ es el índice de enmascaramiento del enmascarador no tonal:

$$av_{nm}(k) = -0.175 \cdot z(k) - 2.025$$

Al igual que en el caso de los enmascaradores tonales, establece un nivel constante, pero en este caso es menor que para los enmascaradores tonales y decrece más rápidamente con la frecuencia medida en Bark.

Umbrales de enmascarmiento individuales



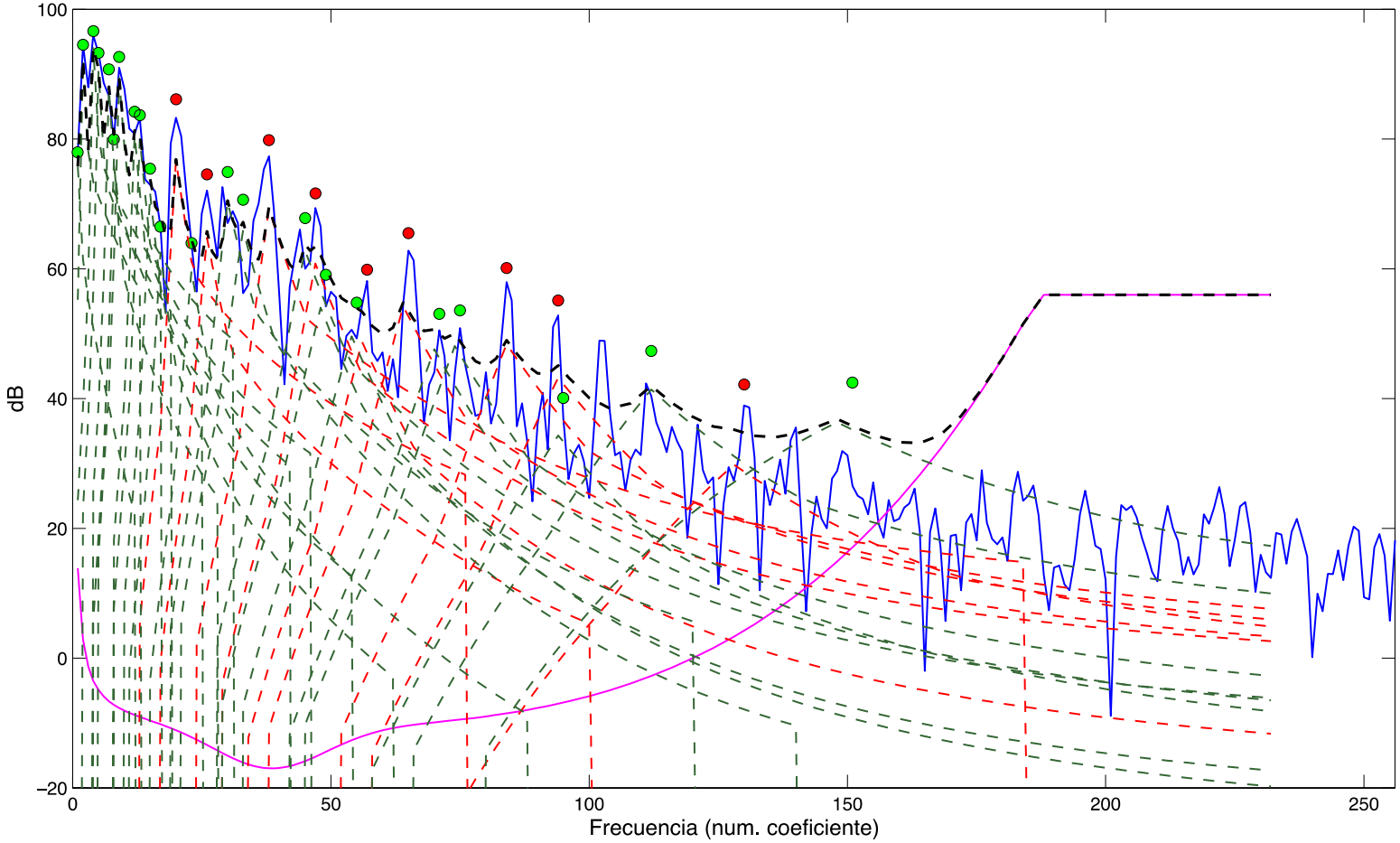
Cálculo del umbral de enmascaramiento global

El umbral de enmascaramiento global se obtiene sumando los umbrales individuales de enmascaramiento originados por cada enmascarador y el umbral absoluto:

$$LT_g(k) = 10\log_{10}\left(10^{\frac{LT_q(k)}{10}} + \sum_{i=1}^{N_m} 10^{\frac{LT_m(i,k)}{10}}\right)$$

donde N_m es el numero total de enmascaradores (tonales y no tonales).

Cálculo del umbral de enmascaramiento global



Cálculo del mínimo umbral de enmascaramiento

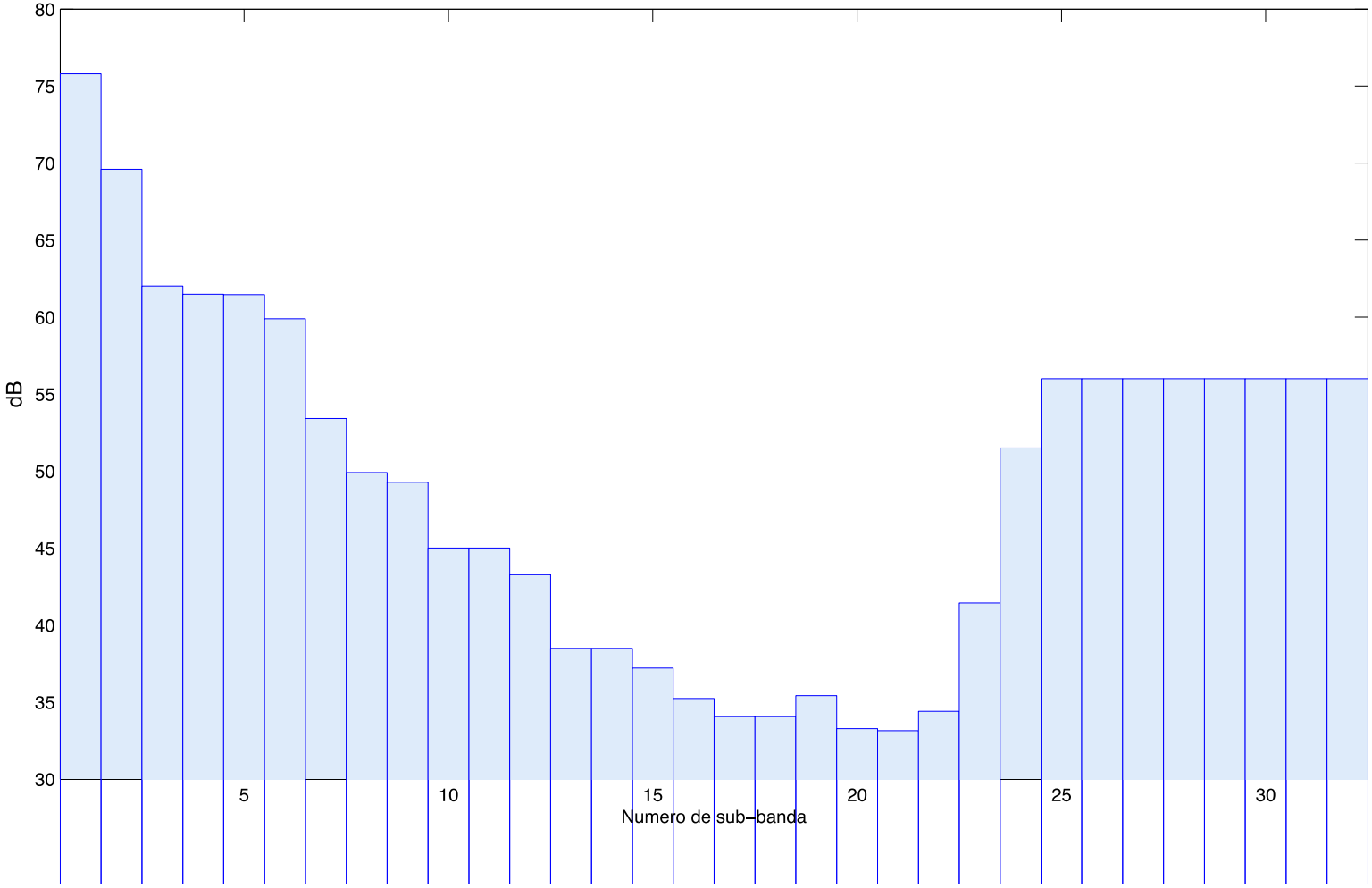
El mínimo umbral de enmascaramiento se calcula encontrando el mínimo del umbral global en cada sub-banda:

$$LT_{\min}(n) = \min_{\substack{\text{k en la n-esima} \\ \text{sub-banda}}} \left(LT_g(k) \right)$$

La relación señal a máscara se calcula como la diferencia entre el nivel de presión sonora para cada sub-banda y el umbral mínimo de enmascaramiento.

$$SMR(n) = L_{sb}(n) - LT_{\min}(n)$$

Mínimo umbral de enmascaramiento



MPEG-1 Audio, Capa I

- ★ Banco de filtros de análisis
- ★ Calculo del factor de escala
- ★ Asignación de bits
- ★ Codificación de la asignación de bits
- ★ Cuantificación y codificación de las muestras de las sub-bandas
- ★ Formateado

Capa I. Banco de filtros de análisis

- El banco de filtros de análisis está basado en la realización eficiente del banco de filtros modulados en coseno propuesta por Joseph Rothweiler en 1983.

- El banco de filtros tienes las siguientes limitaciones:

  - Las respuestas en frecuencia de filtros adyacentes se solapan, por lo que una única frecuencia puede originar salidas en dos filtros adyacentes

  - Los filtros no producen una reconstrucción perfecta por lo que, incluso en ausencia de cuantificación, la señal reconstruida será diferente de la señal de entrada. En cualquier caso la diferencia es tan pequeña que es inaudible.

  - Las sub-bandas tienen todas la misma anchura, por lo que no se corresponden con las características del sistema auditivo humano (bandas críticas de diferente anchura), especialmente a bajas frecuencias.

Capa I. Cálculo del factor de escala

- Se calcula un factor de escala por cada conjunto de 12 muestras de cada filtro de análisis.
- El factor de escala se usa en
  - El cálculo de la relación señal a máscara
  - En la cuantificación y codificación de las muestras de cada sub-banda
- El factor de escala se calcula para cada sub-banda. Se busca la muestra con mayor absoluto de las 12 muestras de la sub-banda y se escoge el menor valor de la tabla que es mayor que el valor máximo. El índice obtenido se transmite con 6 bits (MSB primero) si se ha asignado a la banda un número de bits distinto de cero.

0	2,00000000000000	16	0,04960628287401	32	0,00123039165029	48	0,00003051757813
1	1,58740105196820	17	0,03937253280921	33	0,00097656250000	49	0,00002422181781
2	1,25992104989487	18	0,03125000000000	34	0,00077509816991	50	0,00001922486954
3	1,00000000000000	19	0,02480314143700	35	0,00061519582514	51	0,00001525878906
4	0,79370052598410	20	0,01968626640461	36	0,00048828125000	52	0,00001211090890
5	0,62996052494744	21	0,01562500000000	37	0,00038754908495	53	0,00000961243477
6	0,50000000000000	22	0,01240157071850	38	0,00030759791257	54	0,00000762939453
7	0,39685026299205	23	0,00984313320230	39	0,00024414062500	55	0,00000605545445
8	0,31498026247372	24	0,00781250000000	40	0,00019377454248	56	0,00000480621738
9	0,25000000000000	25	0,00620078535925	41	0,00015379895629	57	0,00000381469727
10	0,19842513149602	26	0,00492156660115	42	0,00012207031250	58	0,00000302772723
11	0,15749013123686	27	0,00390625000000	43	0,00009688727124	59	0,00000240310869
12	0,12500000000000	28	0,00310039267963	44	0,00007689947814	60	0,00000190734863
13	0,09921256574801	29	0,00246078330058	45	0,00006103515625	61	0,00000151386361
14	0,07874506561843	30	0,00195312500000	46	0,00004844363562	62	0,00000120155435
15	0,06250000000000	31	0,00155019633981	47	0,00003844973907		

Capa I. Asignación dinámica de bits

- Cada grupo de 384 muestras es codificado en una trama, compuesta de un número entero de slots de 32 bits.
- El número de bits disponibles se calcula restando al número total de bits disponibles, bc:
  - el número de bits de la cabecera ***bhdr*** (32 bits)
  - El ***CRC***, si se usa ***bcrc*** (16 bits)
  - El contador de bits asignados a cada sub-banda, ***bbal***.
  - Los datos auxiliares, si existen (***banc***)

bits	SN (dB)	A	B
2	7,00	0,750000000	-0,250000000
3	16,00	0,875000000	-0,125000000
4	25,28	0,937500000	-0,062500000
5	31,59	0,968750000	-0,031250000
6	37,75	0,984375000	-0,015625000
7	43,84	0,992187500	-0,007812500
8	49,89	0,996093750	-0,003906250
9	55,93	0,998046875	-0,001953125
10	61,96	0,999023438	-0,000976563
11	67,98	0,999511719	-0,000488281
12	74,01	0,999755859	-0,000244141
13	80,03	0,999877930	-0,000122070
14	86,05	0,999938965	-0,000061035
15	92,01	0,999969482	-0,000030518

- El número de bits disponibles,

  bc-(bhdr+bcrc+bbal+banc)

  se distribuye entre las sub-bandas intentando minimizar la relación ruido a máscara para esa trama.
- El número de bits asignados a una muestra varía entre 0 y 15, con la excepción de 1 que está prohibido. Esto es, puede asignarse 0, 2, 3…15 bits por muestra.

Capa I. Asignación dinámica de bits

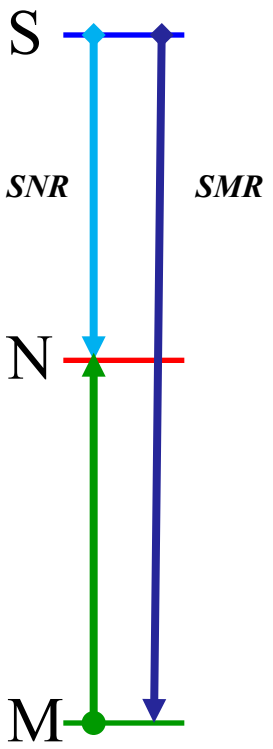
- ★ El procedimiento para asignar el número de bits a cada sub-banda comienza calculando la relación ruido a umbral de enmascaramiento (NMR) para cada sub-banda

  MNR=SNR-SMR

- ★ La SMR se ha obtenido a partir del modelo psicoacústico, mientras que la SNR depende del número de bits asignados a cada sub-banda.
- ★ El procedimiento de asignación es iterativo:
  1. Usar el número de bits asignados a cada sub-banda para calcular la SNR y de ella la MNR
  2. Encontrar la sub-banda con el MNR mínimo (ruido audible si MNR negativo)
  3. Aumentar en uno el número de bits asignados a esa sub-banda.
  4. Recalcular **bspl**, el número de bits necesarios para transmitir las muestras de las sub-bandas. Si el número de bits ha pasado de 0 a 2, además deben añadirse a **bcsf** 6 bits necesarios para transmitir el factor de escala de esa sub-banda.
  5. Recalcular el número de bits disponibles,

     adb=bc-(bhdr+bcrc+bbal+bcsf+bspl+banc)

  6. Si adb es mayor o igual a cero la nueva asignación es posible. Si adb es suficientemente grande puede continuarse con el paso 1.



Capa I. Codificación de la asignación de bits

★ El número de bits asignado a cada sub-banda de cada canal es transmitido usando 4 bits según la tabla adjunta

código	bits/m	código	bits/m	código	bits/m	código	bits/m
0000	**0**	0100	**5**	1000	**9**	1100	**13**
0001	**2**	0101	**6**	1001	**10**	1101	**14**
0010	**3**	0110	**7**	1010	**11**	1110	**15**
0011	**4**	0111	**8**	1011	**12**	1111	**prohibido**

Cuantificación y codificación de las muestras de las sub-bandas

1. Obtener la muestra normalizada, X, dividiendo su valor por el factor de escala

2. Calcular el valor cuantificado mediante la expresión Q=AX+B, donde los valores de A y B dependen del número de bits asignado a la sub-banda (véase la tabla incluida en “Capa I. Asignación dinámica de bits”)

3. Encontrar Q_N reteniendo los N bits más significativos de Q, donde N es el número de bits asignado a la sub-banda. Q se representa como una fracción binaria en complemento a dos

4. Invertir el bit más significativo de Q_N para evitar que la palabra código contenga sólo 1s

Cuantificación y Codificación. Ejemplo

Supongamos que una sub-banda debe cuantificarse con 3 bits.

- ★ Para 3 bits, A=0.875 y B=-0.125
- ★ Como el factor de escala siempre es mayor que la mayor muestra de la sub-banda, los valores de X están comprendidos entre -1 y 1
- ★ Los valores de Q están comprendidos entre 0.875×(-1)-0.125=-1 y 0.875×1-0.125=0.750
- ★ Al representarse la muestra en complemento a 2 hay 7 posibles intervalos, cada uno de ellos de longitud 1.75/7=0.25

Q (intervalo)	X (intervalo)	N(intervalo)	Q_3	código
-1<Q<-0.75	-1< X < -5/7	-4	011+1=100	000
-0.75≤Q<-0.5	-5/7 ≤ X < -3/7	-3	100+1=101	001
-0.5≤Q<-0.25	-3/7 ≤ X < -1/7	-2	101+1=110	010
-0.25≤Q<0.0	-1/7 ≤ X < 1/7	-1	110+1=111	011
0.0≤Q<0.25	1/7 ≤ X < 3/7	0	000	100
0.25≤Q<0.50	3/7 ≤ X < 5/7	1	001	101
0.50≤Q<0.75	5/7 ≤ X < 1	2	010	110

Datos auxiliares (*ancillary data*)

- ★ La norma permite la inclusión de un número variable de datos auxiliares
- ★ El número de bits correspondiente a los datos auxiliares está incluido en el presupuesto de bits de cada trama y por lo tanto reduce el número de bits disponibles para la transmisión de audio
- ★ En consecuencia, un número excesivo de bits dedicados a la transmisión de datos auxiliares producirá una merma significativa de la calidad del audio
- ★ Además debe evitarse que un grupo de bits mimetice la palabra de sincronización (0xFFF) que se incluye al principio de cada trama. Si no se hace así, la recuperación de la sincronización después de un error puede dificultarse notablemente

Capa I. Formateado

- La salida del codificador de audio se transmite en tramas (*frames*). Cada trama contiene la información suficiente para decodificar 384 muestras de la señal de audio original.

- Las Capas I y II producen un flujo binario de salida de tasa binaria constante. El número de bits del que se dispone para codificar cada trama viene determinado por esta velocidad binaria.

- Cada trama debe contener un número entero de “slots”. En la Capa I un slot son 32bits, y el número de slots en una trama se calcula como:

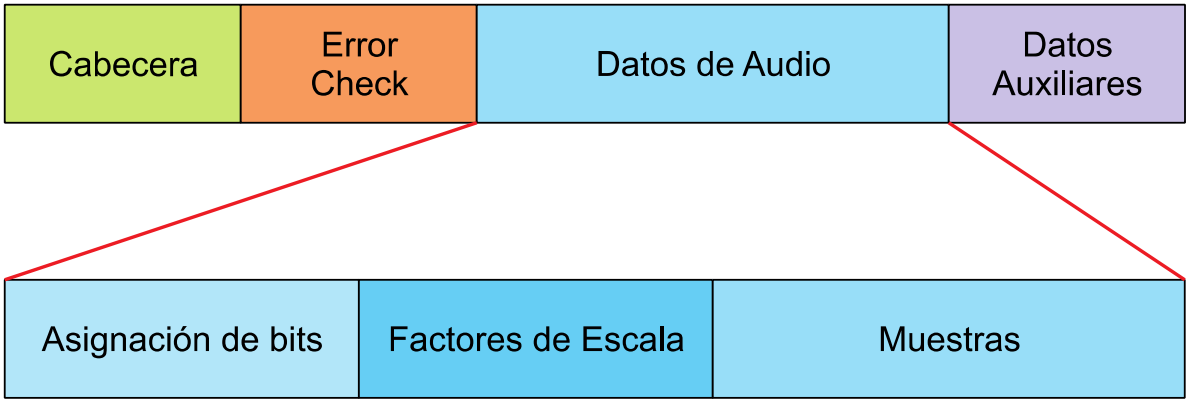
$$N_s = \left\lfloor \frac{384 \cdot R}{32 \cdot F_s} \right\rfloor$$

  donde R es la tasa binaria de salida y F_s la frecuencia de muestreo

- La longitud de una trama debe ser un número entero de slots, de manera que siempre comienza en un límite de slot. Esto se hace para favorecer la resincronización en caso de error.

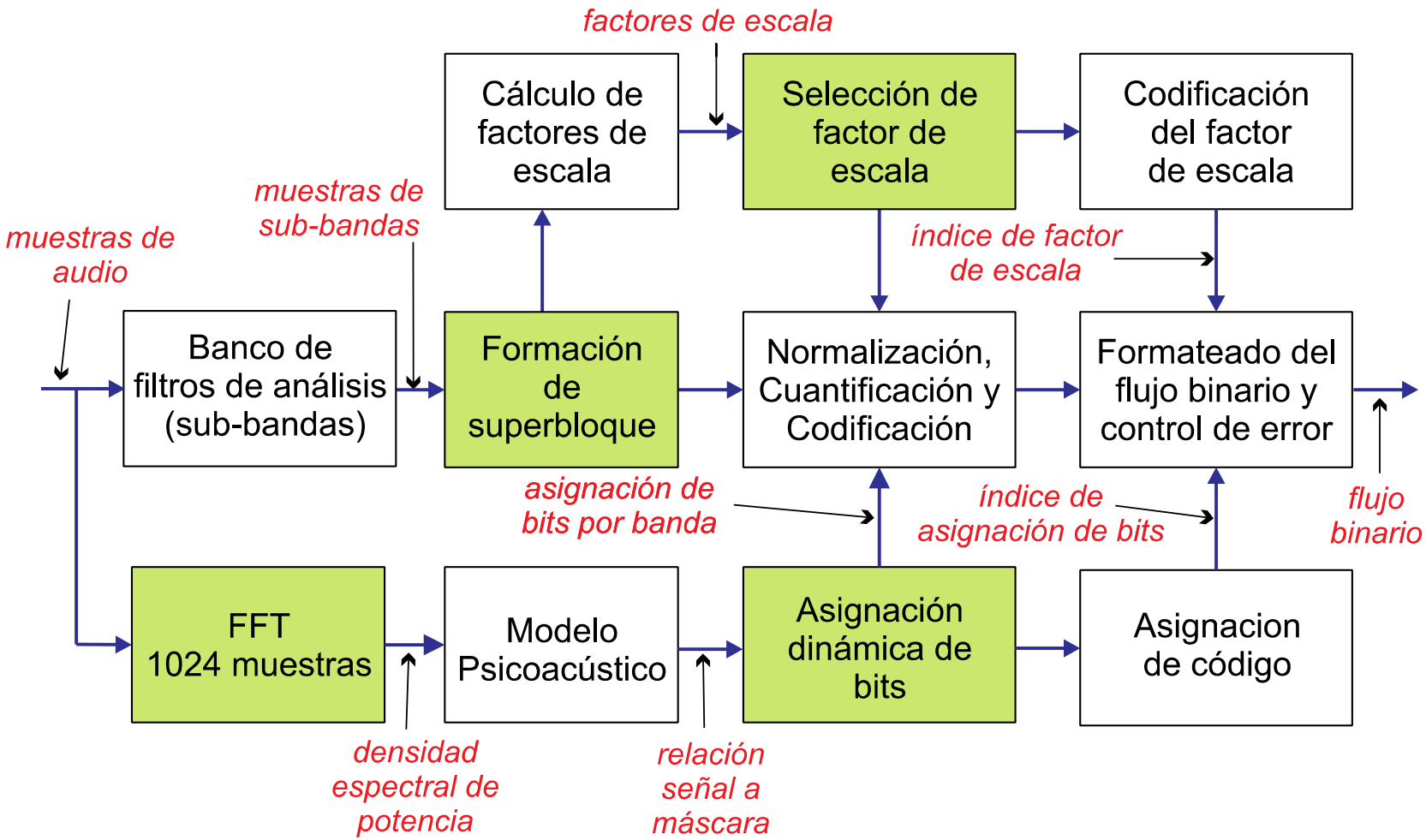
- Para algunas combinaciones de R y F_s , si se mantiene constante Ns, la tasa binaria se sitúa por debajo de R, por lo que en la práctica se permite que el número de slots varíe entre N_s y N_s+1 para mantener constante la tasa binaria.

Capa I: Formato de la trama de audio



MPEG-1 Audio, Capa II

Diagrama de bloques del codificador

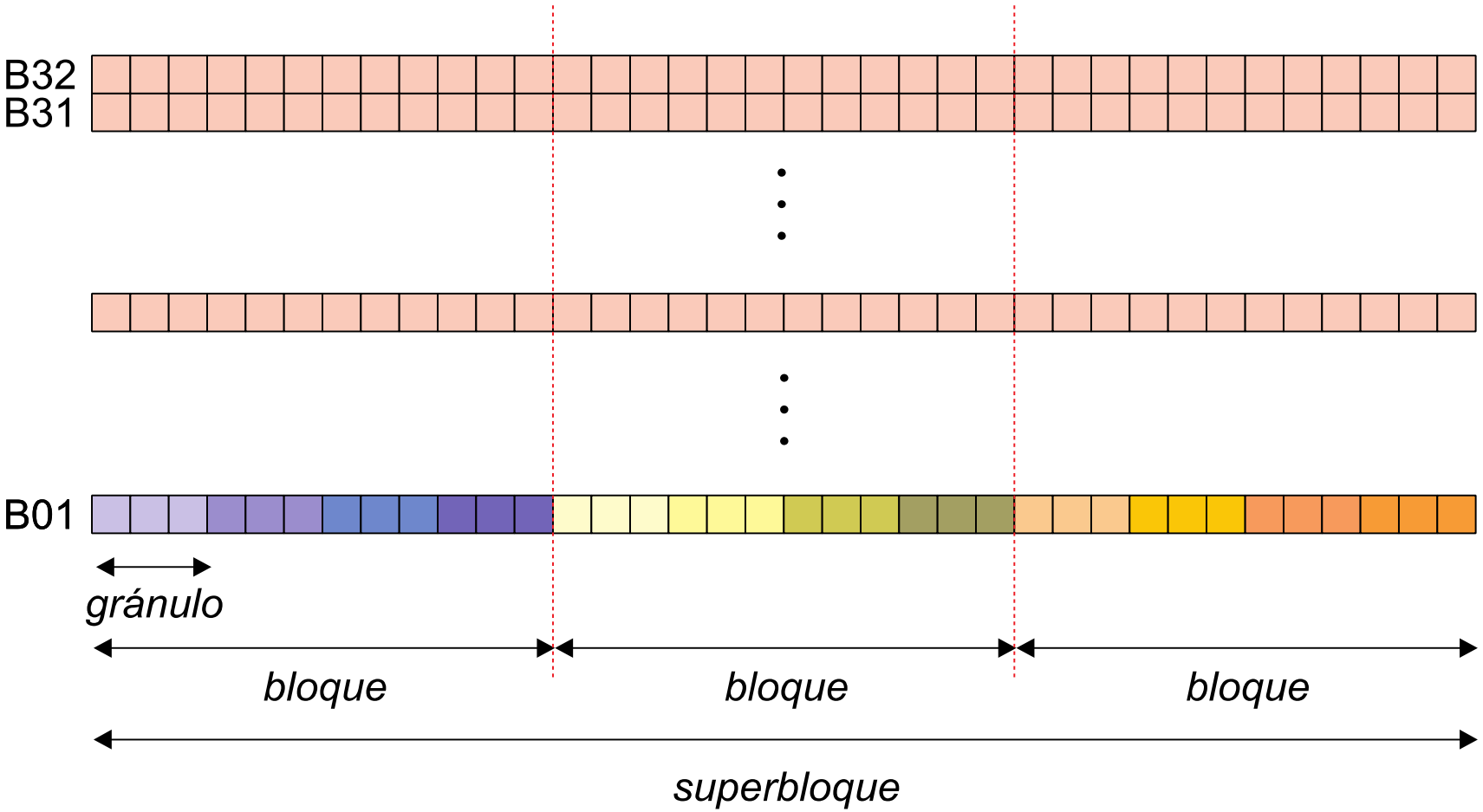


Codificador

Los cambios fundamentales se presentan coloreados en el diagrama.

- ★ Aunque para ambas capas se usa el mismo banco de filtros de análisis, el codificador de la Capa II procesa tres bloques consecutivos de 12x32 muestras (superbloque)
- ★ Para su cuantificación y codificación, cada bloque se considera compuesto de 4 “gránulos” de 3 muestras consecutivas.
- ★ En consecuencia la trama de la Capa II es tres veces mayor (384x3=1152 muestras)
- ★ La FFT se calcula sobre 1024 muestras
- ★ Aunque se sigue calculando un factor de escala para cada bloque de 12 muestras, los tres factores de escala consecutivos de cada superbloque son codificados de forma más compleja, intentando sacar partido de la redundancia existente.
- ★ El número de escalones de cuantificación se reduce en las sub-bandas de frecuencias medias y altas, con la consiguiente ahorro de bits.

Capa II. Muestras de un canal



Capa II. Codificación de los factores de escala

Se calcula un factor de escala para cada bloque de 12 muestras (igual que en la Capa I), luego para cada superbloque deben codificarse 3 factores de escala por sub-banda. Por lo general existirá cierta redundancia entre los valores de los factores de escala, que el codificador trata de explotar con el siguiente procedimiento:

1. Calcular las diferencias entre factores de escala consecutivos:

$dscf_1 = scf_1 - scf_2$ $\qquad$ $dscf_2 = scf_2 - scf_3$

2. Categorizar las diferencias obtenidas en una de las cinco clases siguientes

dscf	clase
$dscf \leq -3$	1
$-3 < dscf < 0$	2
$dscf = 0$	3
$0 < dscf < 3$	4
$dscf \geq 3$	5

Capa II. Codificación de los factores de escala

3. En función de las dos clases obtenidas (25 combinaciones diferentes) se decide:

- Cuántos y cuáles son los factores de escala que van a transmitirse (1, 2 ó 3)
- Qué factor de escala será usado para decodificar cada bloque en el receptor

En las tablas siguientes se indica cuáles son los factores de escala transmitidos para cada una de las 25 combinaciones posibles así como la asignación de los factores de escala transmitidos a cada bloque del superbloque.

Ind. Selección información	Num. scf transmitidos	Asignación
0 (00)	3	sf_1 a bloque 1, sf_2 a bloque 2 y sf_3 a bloque 3
1 (01)	2	sf_1 a bloques 1 y 2 y sf_2 a bloque 3
2 (10)	1	sf_1 a bloques 1, 2 y 3
3 (11)	2	sf_1 a bloque 1 y sf_2 a bloques 2 y 3

Capa II. Codificación de los factores de escala

Clase 1	Clase 2	scf usados	Transmisión	Ind. Selec. Info.	Clase 1	Clase 2	scf usados	Transmisión	Ind. Selec. Info.
1	1	123	123	0	3	4	333	3	2
1	2	122	12	3	3	5	113	13	1
1	3	122	12	3	4	1	222	2	2
1	4	133	13	3	4	2	222	2	2
1	5	123	123	0	4	3	222	2	2
2	1	113	13	1	4	4	333	3	2
2	2	111	1	2	4	5	123	123	0
2	3	111	1	2	5	1	123	123	0
2	4	111	1	2	5	2	122	12	3
2	5	113	13	1	5	3	122	12	3
3	1	111	1	2	5	4	133	13	3
3	2	111	1	2	5	5	123	123	0
3	3	111	1	2					

Capa II. Asignación de bits

- El número total de bits, **cb**, para cada trama de audio se calcula a partir del número de slots (de 8 bits en la Capa II) que van a transmitirse en esa trama.

- El número de bits disponible para transmitir muestras (**adb**) se calcula restando el número de bits correspondientes a
  - la cabecera (**bhdr**)
  - El código CRC ,si se usa (**bcrc**=16)
  - La información de asignación de bits a cada sub-banda (**bbal**)
  - Los datos auxiliares (**banc**)

  adb=cb-(bhdr+bcrc+bbal+banc)

- Los bits se asignan a cada sub-banda minimizando el total “ruido a máscara” para la trama. El número de bits asignado a cada muestra puede variar de 0 a 16 con la excepción de 1, que no puede usarse.

- Como las muestras se codifican en grupos de 3 (“gránulos”), el número de bits asignado a cada grupo de muestras podría ser

  0,6,9, 12,15,18,21,24,27,30,33,36,39,42,45,48

  pero los valores que se usan son:

  0,5,7,9,10,12,15,18,21,24,27,30,33,36,39,42,45,48

Como 5, 7 y 10 no son múltiplos de 3 requieren un tratamiento especial.

Capa II. Asignación de bits

★ En la Capa II se aplican además restricciones adicionales al número de bits que pueden asignarse a cada sub-banda en función de la frecuencia de muestreo y de la velocidad binaria de salida.

★ El procedimiento es

1. Usar el número de bits asignados a cada sub-banda para calcular la SNR y de ella la MNR
2. Encontrar la sub-banda con el NMR mínimo
3. Aumentar el número de bits asignados a esa sub-banda hasta el siguiente valor permitido
4. Recalcular **bspl**, el número de bits necesarios para transmitir las muestras de las sub-bandas. Si el número de bits ha pasado de 0 a 5, además debe añadirse el número de bits necesario para transmitir los factores de escala **bcsf** y la información de selección de factores de escala (**bsel**)
5. Recalcular el número de bits disponibles,

   Adb=cb-(bhdr+bcrc+bbal+bsel+bcsf+bspl+banc)

6. Si adb es mayor o igual a cero la nueva asignación de bits es posible. Si adb es suficientemente grande se vuelve al paso 1.

La asignación de bits por sub-banda se codifica como un entero (MSB *first*) tomado de las tablas de asignación de bits por sub-banda. Para la tabla del ejemplo, en la sub-banda 10, el índice 9 (1001), codificado con 4 bits (nbal=4) indica que el número de bits asignado a cada conjunto de 3 muestras es 24.

Capa II. SNR en función de bits asignados

Núm. bits	SNR(dB)	A	B
0	0		
5	7	0,750000000	-0,250000000
7	11	0,625000000	-0,375000000
9	16	0,875000000	-0,125000000
10	20,84	0,562500000	-0,437500000
12	25,28	0,937500000	-0,062500000
15	31,59	0,968750000	-0,031250000
18	37,75	0,984375000	-0,015625000
21	43,84	0,992187500	-0,007812500

Núm. bits	SNR(dB)	A	B
24	49,89	0,996093750	-0,003906250
27	55,93	0,998046875	-0,001953125
30	61,96	0,999023438	-0,000976563
33	67,98	0,999511719	-0,000488281
36	74,01	0,999755859	-0,000244141
39	80,03	0,999877930	-0,000122070
42	86,05	0,999938965	-0,000061035
45	92,01	0,999969482	-0,000030518
48	98,01	0,999984741	-0,000015259

Capa II. Codificación asignación bits por sub-banda. Ejemplo

sb	nbal	Indice de asignación															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	4		5	9	12	15	18	21	24	27	30	33	36	39	42	45	48
1	4		5	9	12	15	18	21	24	27	30	33	36	39	42	45	48
2	4		5	9	12	15	18	21	24	27	30	33	36	39	42	45	48
3	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
4	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
5	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
6	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
7	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
8	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
9	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
10	4		5	7	9	10	12	15	18	21	24	27	30	33	36	39	48
11	3		5	7	9	10	12	15	48								
12	3		5	1	9	10	12	15	48								

sb	nbal	Indice de asignación															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
13	3		5	7	9	10	12	15	48								
14	3		5	7	9	10	12	15	48								
15	3		5	7	9	10	12	15	48								
16	3		5	7	9	10	12	15	48								
17	3		5	7	9	10	12	15	48								
18	3		5	7	9	10	12	15	48								
19	3		5	7	9	10	12	15	48								
20	3		5	7	9	10	12	15	48								
21	3		5	7	9	10	12	15	48								
22	3		5	7	9	10	12	15	48								
23	2		5	7	48												
24	2		5	7	48												
25	2		5	7	48												
26	2		5	7	48												

Número posible de bits por sub-banda para Fs=32KHz y Fs=44.1 KHz y tasas binarias de 96,112, 128 160 y 192 kbit/s . "nbal" representa el número de bits utilizado para codificar el índice de asignación.

Capa II. Cuantificación y codificación muestras

★ Las muestras de las sub-bandas se codifican con un cuantificador simétrico uniforme con un número impar de intervalos, de manera que los valores próximos a cero se cuantifican como cero.

★ Para cuantificar una muestra:

1. Se normaliza la muestra dividiendo su valor por el factor de escala
2. Se calcula el valor cuantificado mediante la expresión Q=AX+B, donde los valores de A y B dependen del número de bits asignados a la sub-banda (véase la tabla “SNR en función del número de bits asignados”).
3. Obtener Q_N tomando los N bits más significativos de la representación en complemento a 2 de Q.
4. Invertir el bit mas significativo de Q_N para evitar formar una palabra código compuesta exclusivamente por 1s.

★ El codificador de la Capa II transmite grupos de 3 muestras consecutivas. En el caso de que se codifiquen dos canales, los dos conjuntos de tres muestras de cada sub-banda para cada canal se transmiten consecutivamente.

★ Cada muestra se tarta como una palabra código separada excepto si el número de bits asignado es 5, 7 o 10, en cuyo caso las palabras código de cada muestra se tratan como enteros sin signo, se combinan en un único número entero usando las ecuaciones:

$w_5=9w_c+3w_b+w_a$ $w_7=25w_c+5w_b+w_a$ $w_{10}=81w_c+9w_b+w_a$

y se transmite como un entero sin signo de 5, 7 o 10 bits.

Ejemplo

Cuantificar tres muestras consecutivas usando 5 bits.

Dado que el factor de escala es mayor que cualquiera de las muestras de cada sub-banda, el resultado estará comprendido en el intervalo (-1,1). Para 5 bits, A=0,75 y B=-0.25 por lo que:
$Q=0.75X-0.25$, resultando $Q_{max}=0.75\times1-0.25=0.5$ y $Qmin=0.75\times(-1)-0.25=-1$.

En la tabla adjunta se representan los valores posibles de X, Q, Q_2 y la palabra código correspondiente.

De esta forma se producen 3 palabras código posibles que, interpretadas como enteros sin signo, pueden ser 0,1 ó 2. Estas se agrupan para formar una única palabra código mediante la ecuación:

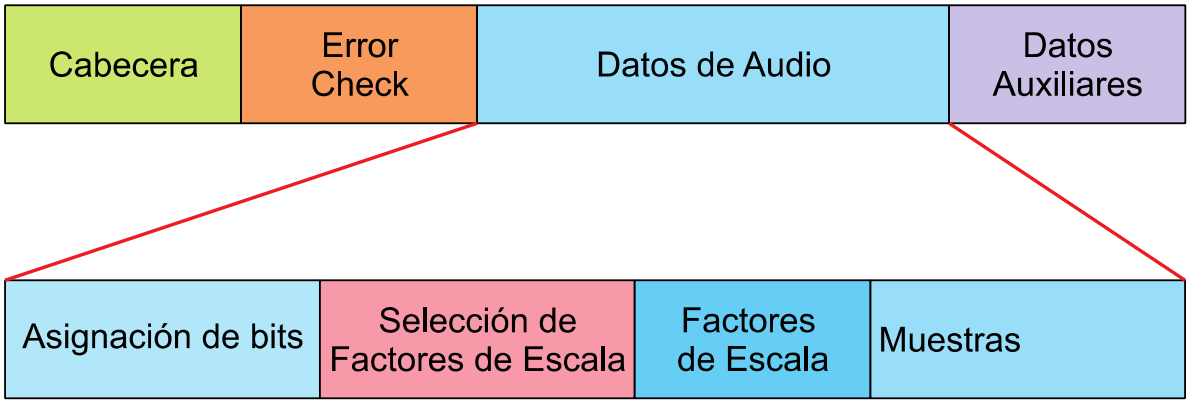
$$w_5=9w_c+3w_b+w_a$$

X	Q	Q_2	Código
-1<X<-⅓	-1<Q<-0.5	-2 (10)	00
-⅓≤X<⅓	-0.5≤Q<0.0	-1 (11)	01
⅓≤X<1	0.0≤Q<0.5	0 (00)	10

que da un número entre 0 y $9\times2+3\times2+2=26$, que se codifica como un entero sin signo de 5 bits (esto es entre 0000 y 11010, ambos inclusive)

Capa II. Formato de la trama de audio

- ★ Las tramas de audio de la Capa II contienen la información necesaria para decodificar 1152 muestras de la señal de audio original
- ★ En la Capa II un slot tiene 8 bits y la trama debe contener un número entero de slots, Ns, que puede calcularse como Ns=int[1152×R/(8Fs)]=int[144R/Fs]
- ★ Como en la Capa I, el número de slots puede variar entre Ns y Ns+1



MPEG-1 Audio, Capa III (MP3)

Codificador Capa III

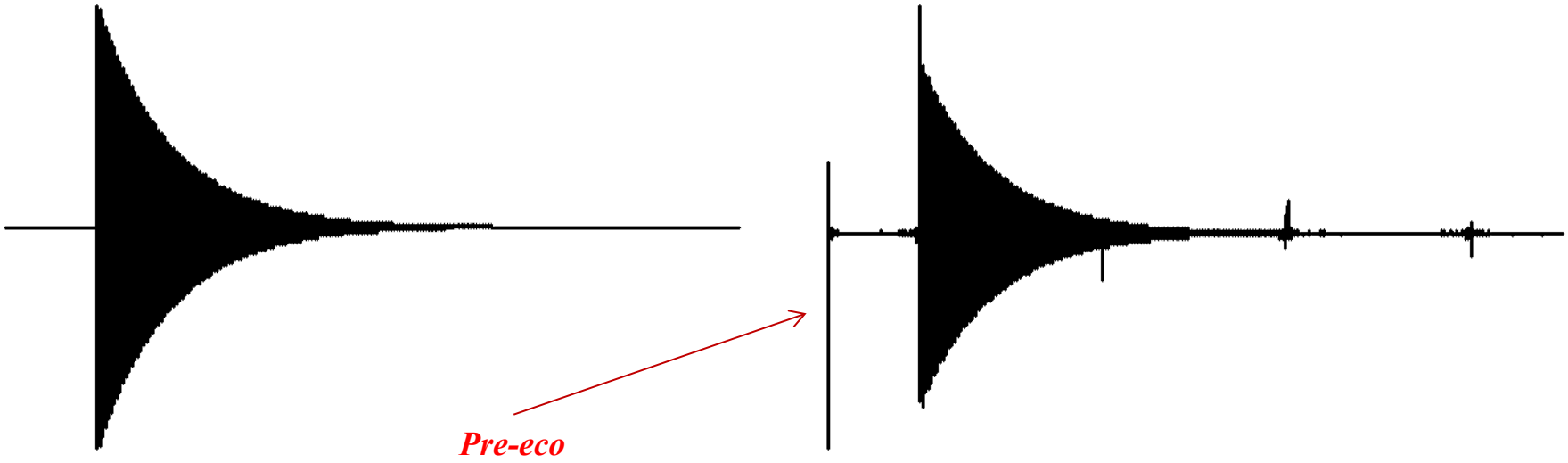
- La Capa III es notablemente más compleja que las Capas I y II.
- Para conseguir una mayor resolución en frecuencia, más próxima a las bandas críticas para frecuencias bajas, se aplica a cada sub-banda una transformada discreta del coseno modificada, con un solapamiento del 50%, de
  - Ventana corta: 6 coeficientes (12 muestras)
  - Ventana larga: 18 coeficientes (36 muestras)
- El número máximo de coeficientes obtenidos es 18×32=576. Para una frecuencia de muestreo de 48KHz, cada coeficiente representa un ancho de banda de tan solo:

$$24000/576=41,67Hz$$

- Los coeficientes de la MDCT son cuantificados y transmitidos al receptor. El error de cuantificación de la DCT se distribuye por todo el bloque, produciendo potencialmente distorsiones audibles (pre-ecos).
- La transformada que se aplica normalmente es la de 18 coeficientes porque proporciona una mayor resolución en frecuencia (a costa de una menor resolución temporal), empleándose la transformada de 6 puntos cuando se esperan pre-ecos (por su mejor resolución temporal).

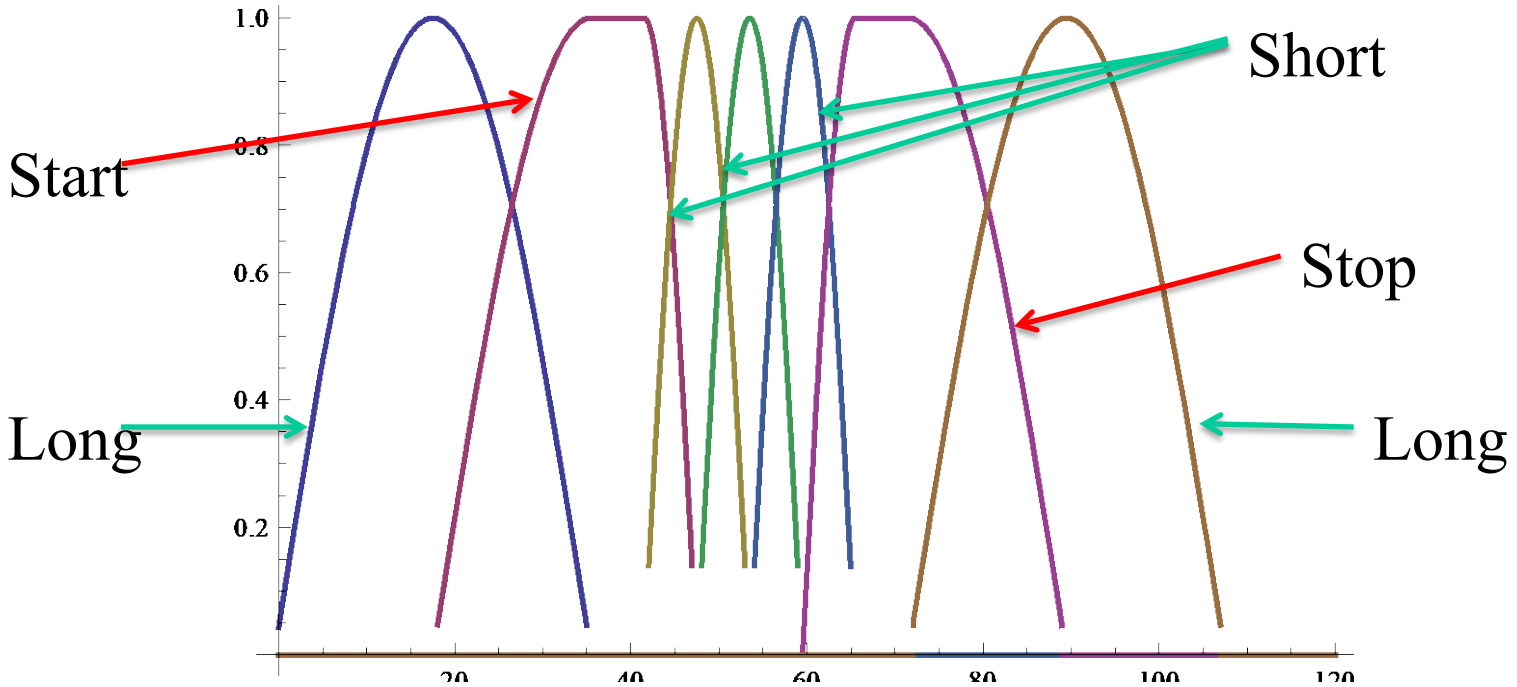
Capa III. Pre-ecos y cambio de ventanas

- ★ Un pre-eco ocurre cuando, después a una señal de nula o muy baja energía le sigue una señal intensa que comienza cerca del final de un bloque. Una situación como esta ocurre por ejemplo cuando un instrumento de percusión como unas castañuelas, o un triángulo, atacan después de un periodo de silencio o casi silencio.
- ★ Por producirse la transformación y cuantificación en bloques, el error de cuantificación se repartirá uniformemente a lo largo de todo el bloque reconstruido.
- ★ Ejemplo: $x(t) = e^{-10(t-0.125)} \cdot u(t-0.125)\sin(4000\pi(t-0.125))$, muestreada a 8KHz. Original y resultado de aplicar la DCT a un bloque de 8192 muestras, cuantificar con 6 bits y calcular la transformada inversa.



Cambio de ventanas

La ventana se escoge en base a la medida de entropía perceptual que proporciona el modelo psicoacústico. Se usa una ventana larga para maximizar la compresión y obtener una buena separación en frecuencia para señales estacionarias. Las ventanas cortas se usan cuando los pre-ecos son probables. El cambio de ventanas largas a cortas y viceversa se realiza utilizando ventanas de transición (ver figura) para preservar la reconstrucción perfecta.



Mejoras de la Capa III frente a las Capas I y II

- **Reducción del solapamiento**. En las ventanas largas se realiza un proceso sobre los coeficientes de la MDCT destinado a reducir algunos efectos derivados del solapamiento de las bandas del banco de filtros de análisis.

- **Cuantificación no uniforme.** La Capa III transforma la señal de entrada según $x^{0.75}$ previamente a su cuantificación uniforme. Este proceso se invierte en el decodificador, pero permite que los valores más grandes sean cuantificados con menor precisión, incrementando la relación señal a ruido para señales de baja intensidad.

- **Factores de escala para bandas**. A diferencia de las Capas I y II, donde cada sub-banda tiene un factor de escala diferente, la Capa III usa bandas cuya anchura se aproxima a las bandas críticas o SFBs (*Scale-Factor Bands*). En la Capa III los factores de escala sirven para colorear el ruido de acuerdo al contorno en frecuencia de los umbrales de enmascaramiento. Los valores de los factores de escala se determinan como parte del proceso de asignación de ruido del codificador.

Codificación Huffman

- Tras la cuantificación el codificador ordena los 576 coeficientes (18 coeficientes MDCT por cada una de las 32 sub-bandas) según un orden predeterminado de frecuencias crecientes, excepto para las ventanas cortas donde se tienen tres valores para una misma frecuencia. En el caso de las ventanas cortas el ordenamiento es primero en frecuencia y luego por ventana dentro de cada “scale-factor band” .

- El codificador delimita a partir de los coeficientes ordenados, comenzando por los coeficientes de alta frecuencia, tres regiones:
  - “rzero”, formada por coeficientes que son todo ceros.
  - “count1”, formada por coeficientes que toman los valores -1, 0 ó +1
  - “bigvalues”, formada por el resto de los coeficientes.

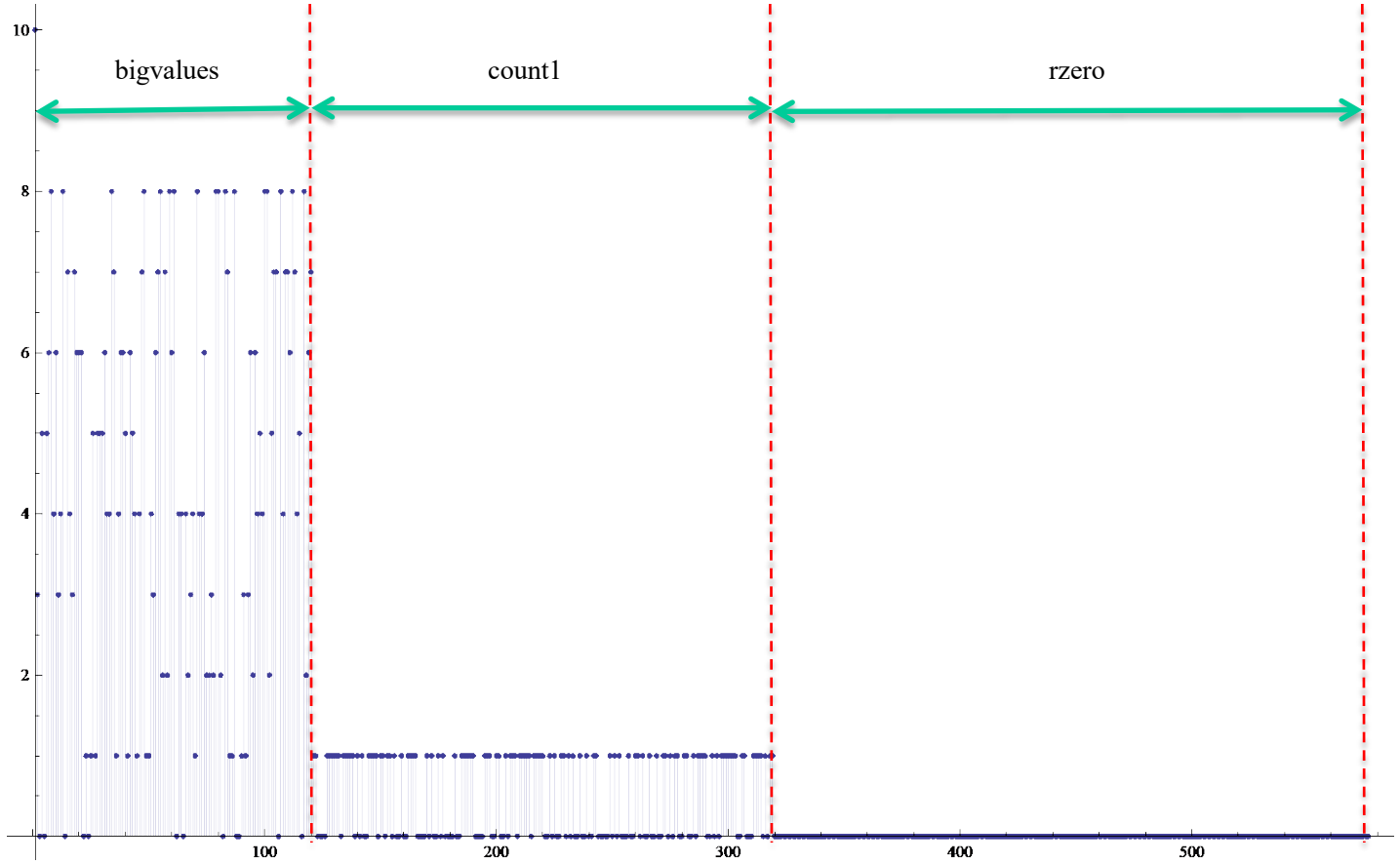
- El codificador usa un conjunto de códigos Huffman diferentes para cada región

- Los valores de la región “bigvalues” se codifican en parejas, si bien se subdivide a su vez en tres regiones, cada una con su código Huffman específico.

- Para la región “count1” se codifican 4valores al tiempo, por lo que su tamaño debe ser múltiplo de 4.

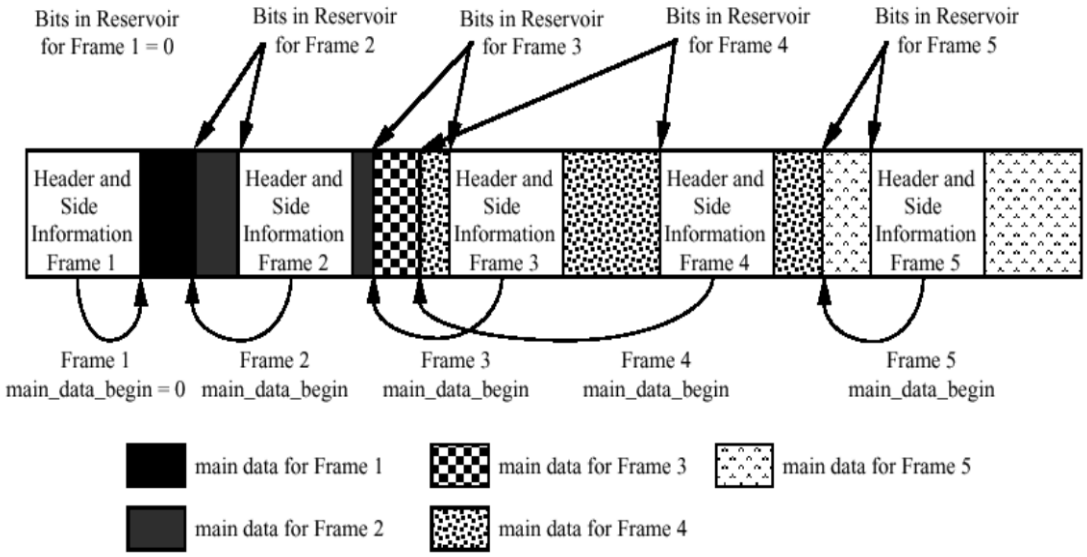
- La región “rzero” no necesita ser codificada puesto que su tamaño puede deducirse del tamaño de las otras regiones (aunque por eso mismo su tamaño debe ser múltiplo de 2)

Codificación Huffman



Bit reservoir

- ★ Al igual que en la capa II, la capa III procesa las muestras de audio en tramas de 1152 muestras, pero a diferencia de la capa II, los datos codificados no se asignan de forma fija a una única trama código. Cuando el codificador necesita menos bits que el valor medio esperado, puede “donar” ese exceso de bits a una reserva. Más adelante, si el codificador necesita un número de bits mayor que la media puede usar esa reserva.
- ★ El codificador puede utilizar solamente bits que hayan sido previamente donados. No puede tomar prestado del futuro.
- ★ El flujo binario de la capa III incluye en la información lateral de cada trama un puntero de 9 bits (main_data_begin) que apunta al comienzo de los datos de esa trama.



Intensity stereo

- Para transmitir una señal estéreo, los canales izquierdo y derecho son codificados de forma separada.
- Sin embargo, si no se dispone de suficientes bits para alcanzar la calidad deseada en ciertas tramas puede utilizarse el modo Intensity_Stereo.
- Este modo de codificación utiliza una característica del sistema auditivo por el cuál, sobre todo a medias y altas frecuencias, la sensación de procedencia del sonido (estéreo) depende no tanto del contenido de los canales izquierdo y derecho sino de su intensidad relativa.
- En el modo Intensity_Stereo, las bandas medias y altas de los canales izquierdo y derecho se suman y el resultado es lo que se cuantifica, codifica y transmite en un único canal, aunque para mantener la sensación estéreo se transmiten los factores de escala (SFn) para cada bloque de cada canal por separado, de manera que la amplitud de los canales izquierdo y derecho puede ser controlada por separado durante la reproducción.
- Un campo de la cabecera de la trama indica el número de la subbanda en la que comienza el modo Intensity_Stereo.

MPEG-2 Audio

MPEG-2 Audio

- MPEG-2 extiende la funcionalidad de MPEG-1 permitiendo menores velocidades de muestreo y permitiendo la codificación de audio multicanal.

- El estándar MPEG-2 contiene 2 algoritmos separados
  - MPEG-2 BC (backward compatible) , diseñado para permitir la compatibilidad con MPEG1. Finalizado en 1995 con una enmienda introducida en 1996
  - MPEG-2 AAC (Advanced Audio Coding) introducido como una nueva parte del estándar MPEG-2 en 1997

- MPEG-2 BC usa los mismos algoritmos de MPEG-1, pero
  - permite menores frecuencias de muestreo para las señales mono y estéreo
  - Soporta audio multicanal a las frecuencias de muestreo de MPEG-1
  - La compatibilidad hacia atrás significa que
    - un decodificador MPEG-2 puede decodificar la señal Audio MPEG -1
    - Un decodificador MPEG-1 puede derivar una señal audio estéreo del flujo binario multicanal MPEG-2 siempre que se hayan tomado las adecuadas precauciones antes de codificar (matrixing). Un codificador MPEG-2 puede producir un flujo binario incompatible con un decodificador MPEG-1

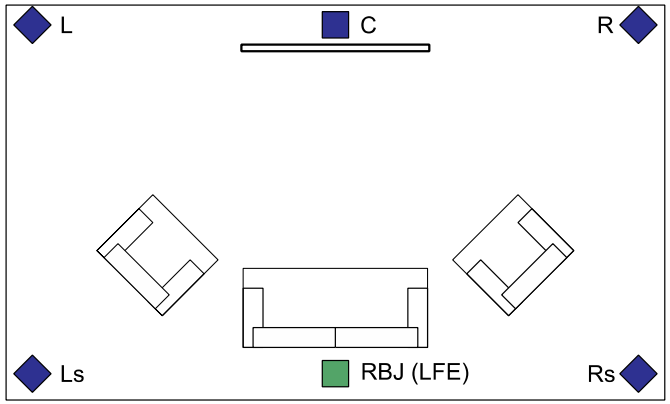
- MPEG-2 AAC se diseñó para incluir los últimos avances en compresión de audio que no podrían realizarse sin romper la compatibilidad con MPEG-1. Permite obtener la misma calidad con una tasa binaria significativamente menor.

Audio multicanal

- Desde hace mucho tiempo se sabe que para producir un campo sonoro realista se requieren al menos cuatro canales, aunque las limitaciones técnicas hicieron que el sonido estéreo de dos canales fuese el estándar aceptado.

- ITU-R recomienda un sistema audio de cinco canales (estéreo 3/2) consistente en los canales izquierdo y derecho (L y R), un canal central (C) y dos canales traseros (L_s y R_s). Este sistema produce un sonido envolvente realista, con una imagen sonora frontal estable y una gran área de audición.

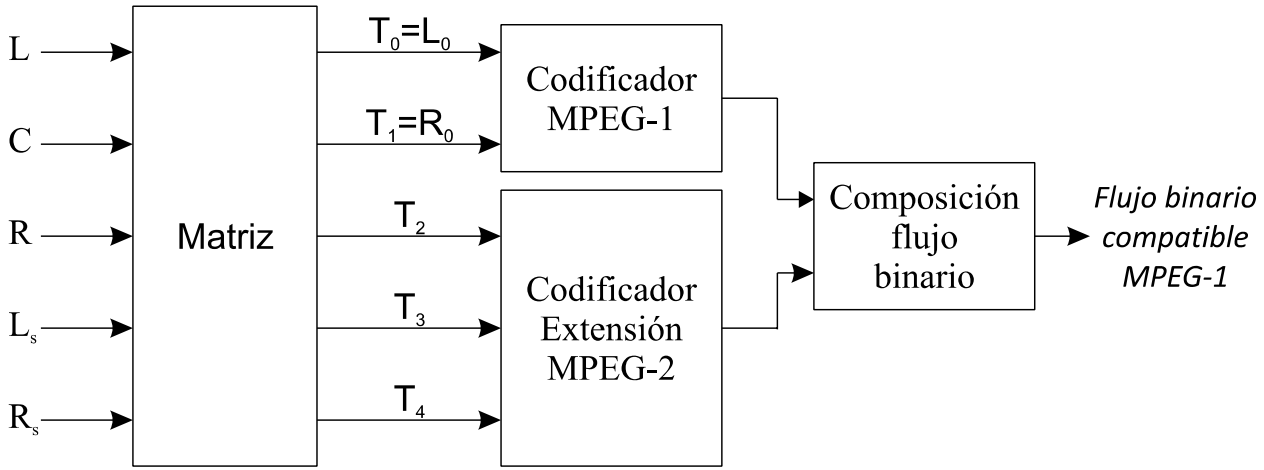
- Puede añadirse opcionalmente un canal opcional de realce de bajas frecuencias (LFE) para incrementar el nivel de las componentes entre 15 y 120Hz.



★ *Un sistema estéreo 3/2 combinado con un canal LFE se conoce habitualmente como 5.1*

Construcción de tramas MPEG-2 Audio BC

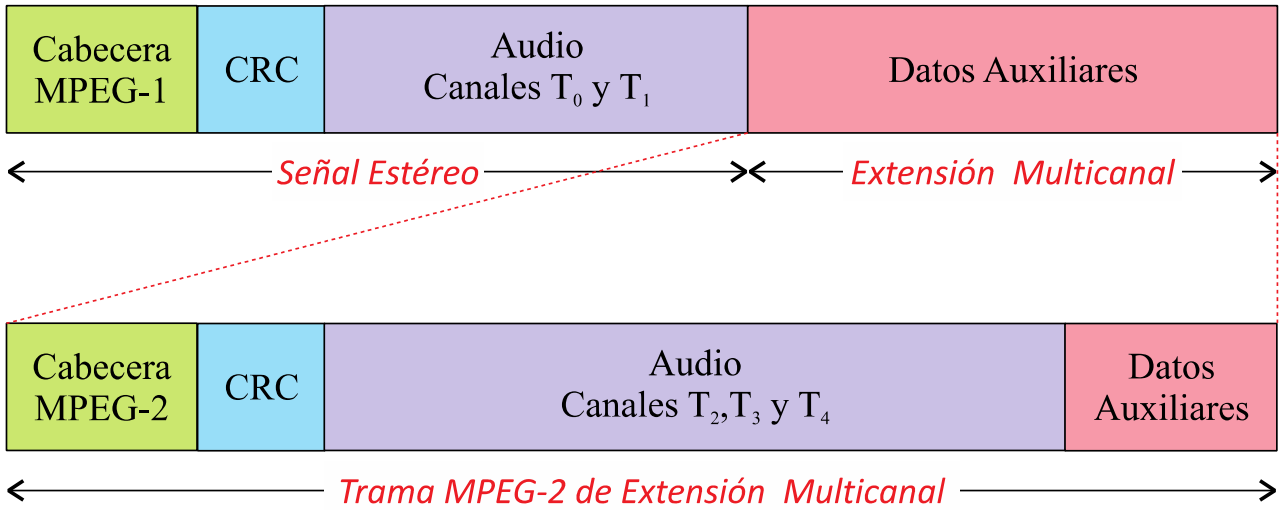
★ Para transmitir una señal de 5.1 canales usando MPEG-2 y compatibilidad hacia atrás el primer proceso a realizar es la formación de una pareja estéreo básica a partir de los canales MPEG-2. Este proceso se conoce como “matrixing”



- Los canales T_0 y T_1 contienen el par estéreo resultante, codificado usando un codificador MPEG-1. Los canales T_2, T_3 y T_4 contienen la información necesaria para reconstruir la señal multicanal original.

Construcción de tramas MPEG-2 Audio BC

★ La compatibilidad hacia atrás se consigue disponiendo los datos correspondientes a T_0 y T_1 en los campos de audio de la trama MPEG-1 , mientras que los datos correspondientes a la extensión multicanal, T_2, T_3 y T_4 viajan en la zona correspondiente a los datos auxiliares MPEG-1. Un codificador MPEG-1 ignorará los datos auxiliares y decodificará el par estéreo L_0/R_0, mientras que un decodificador MPEG-2 recuperará también estos datos y realizará el proceso inverso para recuperar la información multicanal.



Construcción de tramas MPEG-2 Audio BC

La norma MPEG-2 define 4 posibles procedimientos para obtener el par estéreo L_0/R_0 a partir de una señal 5.1. La ecuación utilizada para los procedimientos 0 1 y 3 es:

$$L_0 = \alpha\left(L + \beta C + \gamma L_S\right)$$
$$R_0 = \alpha\left(R + \beta C + \gamma R_S\right)$$

El procedimiento 2 se usa para generar un par estéreo L_0/R_0 compatible con un decodificador Dolby Pro Logic y no lo detallaremos aquí.

Procedimiento	α	β	γ
0	$\frac{1}{1+\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
1	$\frac{1}{1.5+0.5\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$
3	1	0	0

AAC (*Advanced Audio Coding*)

en MPEG-2 y MPEG-4

MPEG-2 AAC

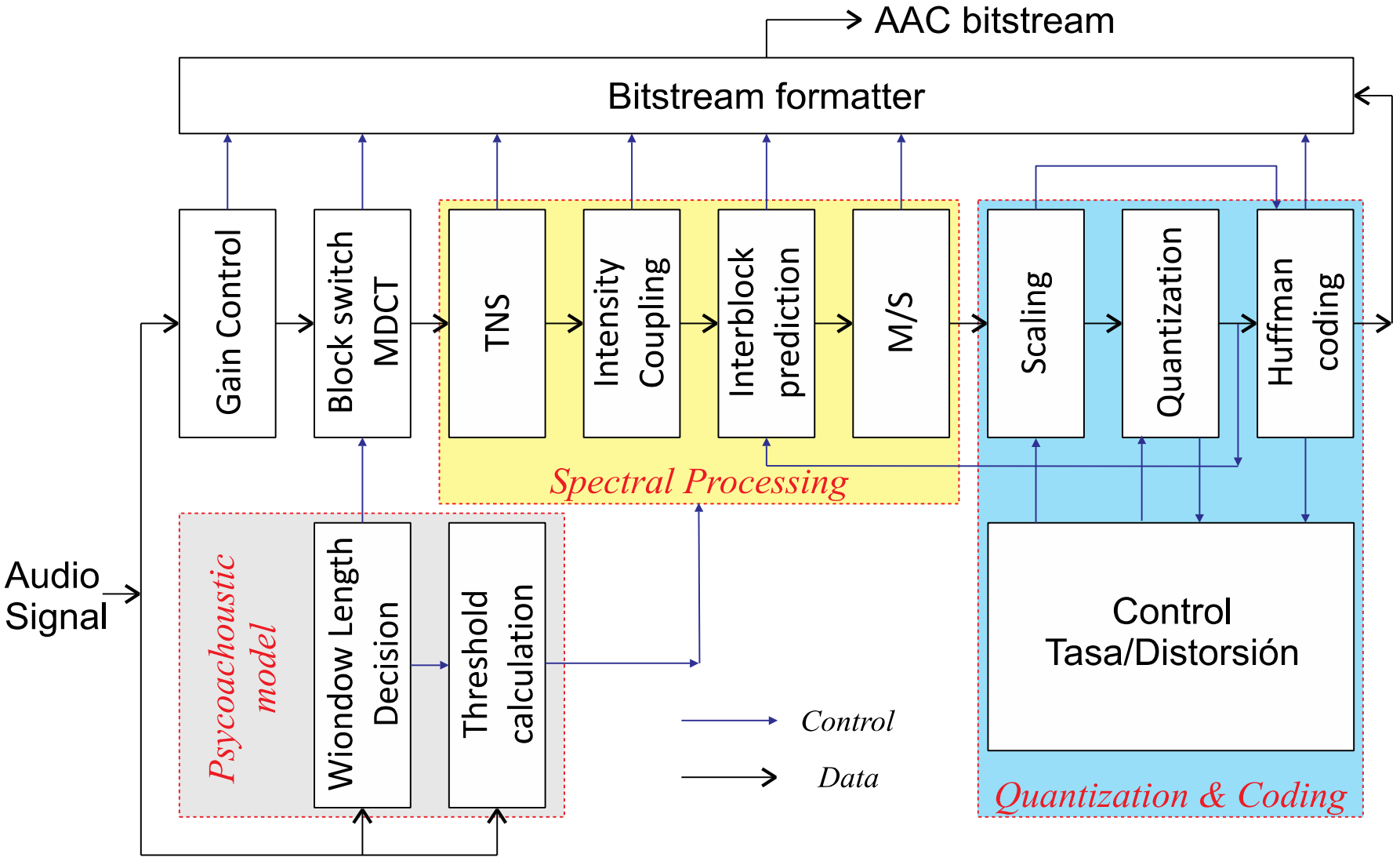
★ El algoritmo de la capa III de MPEG-1 tiene ciertas limitaciones originadas por las restricciones de diseño impuestas, la principal de ellas, la compatibilidad hacia atrás. Este es el motivo de la existencia de una descomposición en 32 sub-bandas seguida de una MDCT.

★ El estándar AAC representa un alternativa multicanal de mayor calidad que MPEG-2 BC. La norma especifica un conjunto de herramientas (“tools”). Como se ilustra en la tabla adjunta, algunas herramientas son necesarias para todos los perfiles(“profiles”) mientras que otras son requeridas sólo por ciertos niveles.

Tool Name	
Bitstream Formatter	Required
Huffman Decoding	Required
Inverse Quantization	Required
Rescaling	Required
M/S	Optional
Interblock Prediction	Optional

Tool Name	
Intensity	Optional
Depend. Switch. Coup.	Optional
TNS	Optional
Block Switching/MDCT	Required
Gain Control	Optional
Indep. Switch. Coup.	Optional

Esquema AAC


AAC bitstream
Bitstream formatter
Gain Control
Block switch MDCT
TNS
Intensity Coupling
Interblock prediction
M/S
Scaling
Quantization
Huffman coding
Spectral Processing
Audio Signal
Psycoachoustic model
Wiondow Length Decision
Threshold calculation
Control
Tasa/Distorsión
Control
Data
Quantization & Coding

Bibliografía

- M. Bosi and R.E. Goldberg, "***Introduction to Digital Audio Coding and Standards***", Kluwer Academic Publishers, 2003.
- Y. You, "***Audio Coding. Theory and Applications***", Springer, New York, 2010.
- J.Arnold, M. Frater and M.Pickering, ***"Digital Television. Technology and Standards***", Wiley-Interscience, 2007.
- B.G.Haskell, A. Puri and A.N. Netravali, ***"Digital Video: An introduction to MPEG-2***", Kluwer Academic Publishers, 1997.
- J.J.Thiagarajan and A.Spanias, "***Analysis of the MPEG-1 Layer III (MP3) Algorithm Using MATLAB***",Synthesis Lectures on Algorithm and Software Engineering, Morgan & Claypool Publishers, 2012.
- F.Pereira and T.Ebrahimi (editores), "***The MPEG-4 Book***", Prentice Hall, New Jersey, 2002.
- P. Noll, "**MPEG Digital Audio Coding**", IEEE Signal Processing Magazine, September 1997, pp 59-81.
- J.Herre and M. Dietz, "**MPEG-4 High Efficiency AAC Coding**", IEEE Signal Processing Magazine, pp. 137-142.
- D.Pan, "**A tutorial on MPEG/Audio Compression**", IEEE Multimedia, Summer 1995.