

# Percepción y modelos psicoacústicos

## 3.1. Introducción

En este documento se proponen algunos ejercicios sobre percepción y modelos psicoacústicos que complementan a los ya propuestos en los apuntes de clase. Todos ellos conformarán esta práctica.

## 3.2. Umbral auditivo absoluto

**Ejercicio 1** Utilizando “*umbralau.m*” representa tu umbral auditivo absoluto. Compáralo con la aproximación:

$$T_q(f) = 3,64(f/1000)^{-0,8} - 6,5e^{-0,6(f/1000-3,3)^2} + 10^{-3}(f/1000)^4 \quad (3.1)$$

## 3.3. Enmascaramiento

**Ejercicio 2** Entre los distintos tipos de enmascaramiento son de especial interdesde el punto de vista de la codificación de audio los casos de TMN y NMT ilustrados en la figura 3.1. Con la ayuda de “*enmas\_tonoruido.m*” comprueba la veracidad de los ejemplos de dicha figura para distintas bandas críticas.

**Ejercicio 3** Hemos visto en clase, con la ayuda de la figura 3.2, que el efecto del enmascaramiento no se limita al interior de una banda crítica, sino que se extiende a las adyacentes. En “*enmas\_tonotono.m*” se considera el caso en que tanto la señal enmascaradora como la señal a enmascarar son dos tonos. Obtén de forma aproximada el umbral de enmascaramiento para distintas frecuencias en las proximidades del tono enmascarador.

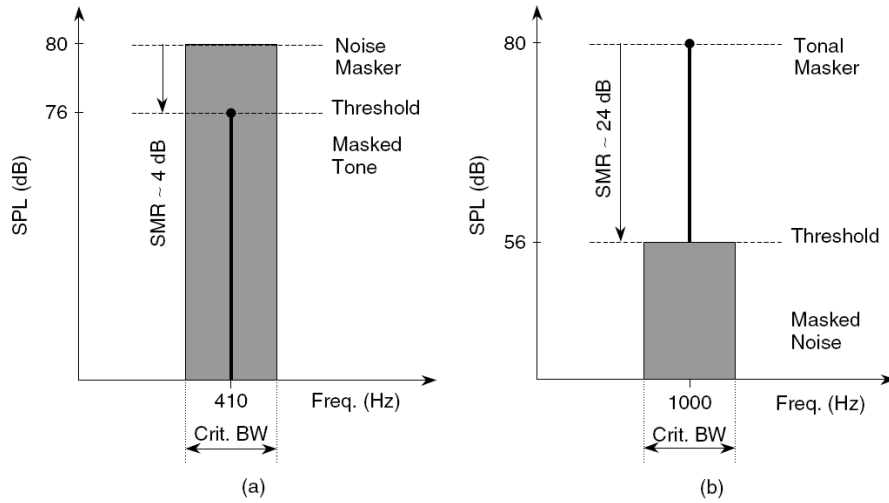


Figura 3.1:

### 3.4. Ejemplo de modelo psicoacústico: Modelo psicoacústico 1 del ISO/IEC 11172-3 (MPEG-1)

En [Painter and Spanias, 2000] y [Spanias et al., 2007] se realiza una descripción mucho más detallada de este modelo psicoacústico.

#### 3.4.1. Paso 1: Normalización y análisis espectral

Dada la señal de audio  $s(n)$  y el número de bits por muestra  $b$ , normalizamos  $s(n)$  de forma que su máxima amplitud sea 1, es decir,

$$x[n] = \frac{s[n]}{2^{b-1}} \quad (3.2)$$

La señal normalizada  $x(n)$  es analizada por tramas de 12 ms (512 muestras), enventanadas con una ventana Hanning, y con un solapamiento entre tramas de 1.1 ms.

A continuación se estima la densidad espectral de potencia,  $P(k)$ , utilizándose una FFT de 512 puntos (la longitud de la trama analizada), mediante

$$P(k) = PN + 10 \log_{10} \left| \frac{1}{N} \sum_{n=0}^{N-1} w(n)x(n)e^{-j2\pi kn/N} \right|^2 \quad 0 \leq k \leq N/2 \quad (3.3)$$

siendo  $PN = 90,3 \text{ dB}$  ( $10 \log_{10}(2^{(b-1)^2})$ ) un término de normalización de potencia. Con este procedimiento se puede considerar que  $P(k)$  es una estimación razonable del nivel de presión sonora (SPL).

#### 3.4.2. Paso 2: Identificación de las componentes enmascaradoras tonales y de ruido.

- Se consideran componentes tonales los máximos locales de  $P(k)$  que superan en más de 7 dB a las componentes espectrales vecinas situadas dentro de una distancia Bark

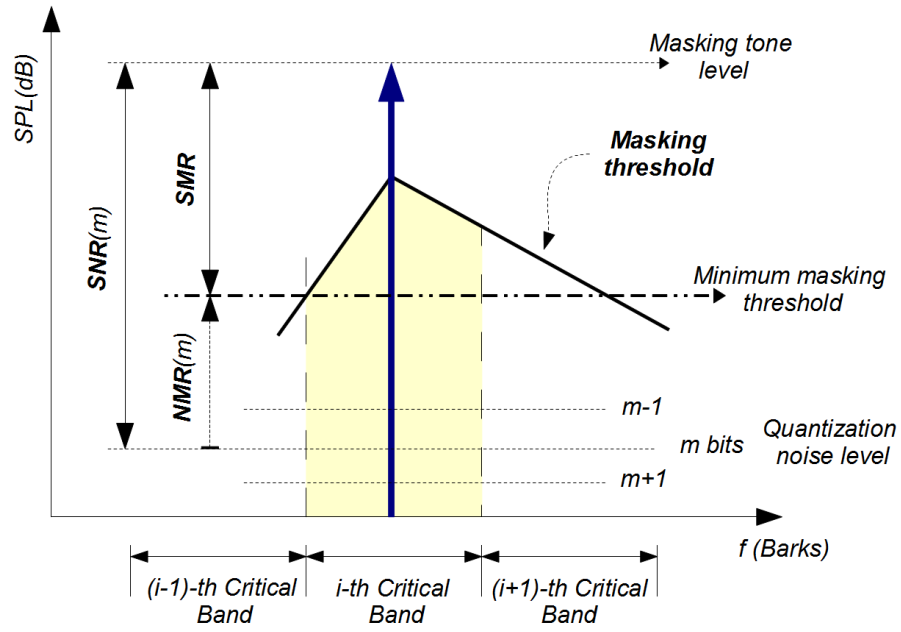


Figura 3.2:

determinada. Así se define el conjunto tonal  $S_T$  como

$$S_T = \left\{ P(k) \mid \begin{array}{l} P(k) > P(k \pm 1) \\ P(k) > P(k \pm \Delta_k) + 7 \end{array} \right\} \quad (3.4)$$

siendo

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & (0,17 - 5,5kHz) \\ [2, 3] & 63 \leq k < 127 & (5,5 - 11kHz) \\ [2, 6] & 127 \leq k \leq 256 & (11 - 20kHz) \end{cases} \quad (3.5)$$

- Los enmascaradores tonales se obtienen añadiendo a la densidad espectral de potencia de cada uno de los picos del conjunto tonal las contribuciones de las componentes espectrales situadas inmediatamente a su izquierda y derecha, es decir,

$$P_{TM}(k) = 10 \log_{10} \left[ \sum_{j=-1}^1 10^{0,1P(k+j)} \right] (dB) \quad (3.6)$$

- A continuación se calcula una componente de ruido enmascaradora para cada banda crítica, utilizando las componentes espectrales  $P(k)$  que no están en la vecindad  $\pm \Delta_k$  de un enmascarador tonal.

$$P_{NM}(k) = 10 \log_{10} \left[ \sum_j 10^{0,1P(j)} \right] (dB) \quad (3.7)$$

### 3.4.3. Paso 3: Diezmado de las componentes enmascaradoras.

- Cualquier componente tonal o de ruido por debajo del umbral auditivo absoluto,  $T_q(k)$ , se descarta.
- Si hay dos o más componentes situadas a menos de 0.5 barks nos quedamos únicamente con la de mayor potencia.

### 3.4.4. Paso 4: Cálculo de los umbrales de enmascaramiento individuales.

Se tiene en cuenta la extensión en frecuencia del enmascaramiento de forma similar a la ilustrada en la figura 3.2.

Así se define  $T_{TM}(i, j)$  que refleja la contribución de la componente tonal situada en la frecuencia  $j$  al umbral de enmascaramiento en la frecuencia  $i$ . Análogamente  $T_{NM}(i, j)$  representa las contribuciones de las componentes de ruido.

### 3.4.5. Paso 5: Cálculo de umbral global de enmascaramiento

Se calcula mediante la suma de todas las contribuciones, es decir,

$$T_g(i) = 10^{0,1T_q(i)} + \sum_{l=1}^L 10^{0,1T_{TM}(i,l)} + \sum_{m=1}^M 10^{0,1T_{NM}(i,m)} \quad (3.8)$$

**Ejercicio 4** Utilizando el código Matlab<sup>1</sup> del capítulo 5 del libro [Spanias et al., 2007], comprueba paso a paso el funcionamiento del modelo psicoacústico, analizando diversos archivos de audio.

<sup>1</sup>El código de todos los capítulos está disponible mediante ftp anónimo en <ftp://ftp.wiley.com/public/sci-tech-med/audio-signal>

# Bibliografía

[Painter and Spanias, 2000] Painter, T. and Spanias, A. (2000). Perceptual coding of digital audio. , 88(4):449-513. abril 2000. *Proceedings of the IEEE*, 88(4):449–513.

[Spanias et al., 2007] Spanias, A., Painter, T., and Attii, V. (2007). *Audio Signal Processing and Coding*. Wiley-Interscience.