Psychoacoustics

Digital Audio Coding

sebastien.boisgerault@mines-paristech.fr

Acoustics

P : quadratic mean (RMS) **sound pressure**

$$P^2 = \langle p^2 \rangle = \frac{1}{N} \sum_{n=0}^{N-1} p(t)^2 \, dt$$

L : **sound pressure level** (dB).

$$L = 10 \log_{10} \frac{P^2}{P_0^2}, \; P_0 = 20 \, \mu\mathrm{Pa}$$

I : **sound intensity**

$$\frac{I}{I_0} = \frac{P^2}{P_0^2}, \; I_0 = 10^{-12} \, \mathrm{W/m^2}$$

Acoustics

$$P^2 = \frac{2}{N\Delta t}\int_0^{\Delta f/2} |p(f)|^2\, df$$

$\ell(f)$: **sound (intensity) density** (dB)

$$\ell(f) = 10\log_{10}\frac{2}{N\Delta t}\frac{|p(f)|^2}{P_0^2}$$

$$10^{L/10} = \int_0^{\Delta f/2} 10^{\ell(f)/10}\, df$$

Acoustics

Normalized sound pressure waveforms

$$x(t) \in [-1.0, +1.0]$$

are interpreted as:

$$p(t) = 10^{4.8} P_0 \times x(t)$$

This convention yields:

$$L = 10 \log_{10} \left\langle x^2 \right\rangle + 96\,\mathrm{dB}$$

$$\ell(f) = 10 \log_{10} \left(\frac{2}{N \Delta t} |x(f)|^2 \right) + 96\,\mathrm{dB}$$

Dynamic Range - 16 bit

Maximal SPL:

$$|x(t)| = 1 \quad \rightarrow \quad L = 96 \text{ dB}$$

"Minimal" SPL : Quantization Noise SPL

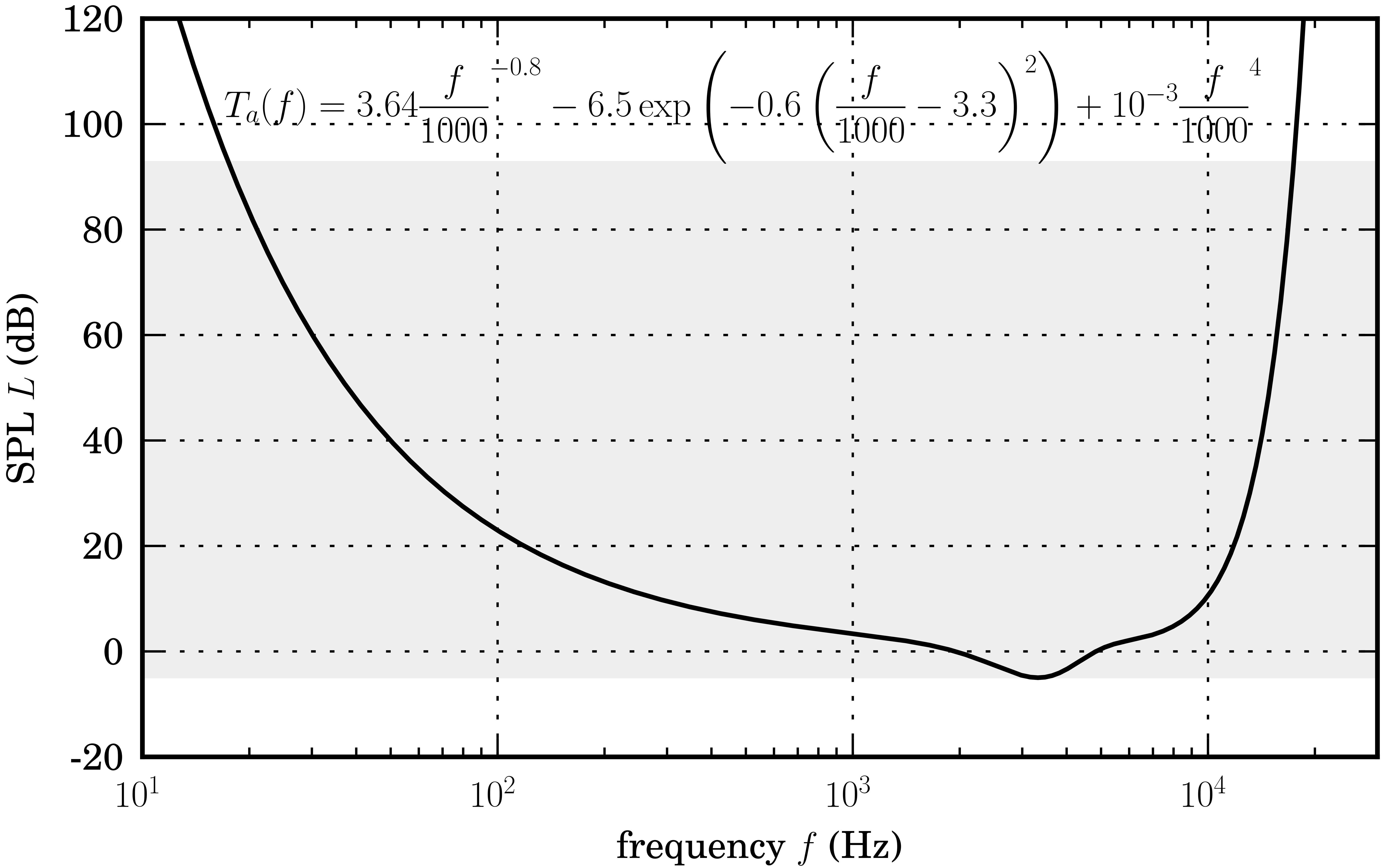
$$\mathbb{E}(X - [X])^2 \simeq \frac{1}{12}\mathbb{E}\Delta(X)^2 = \frac{1}{12}\left(\frac{2}{2^{16}}\right)^2$$

$$L = 10\log_{10}\mathbb{E}(X - [X])^2 + 96 \text{ dB} \simeq -5.1 \text{ dB}$$

Dynamic Range: max SPL - min SPL

$$96 - (-5.1) \simeq 100 \text{ dB}.$$

Absolute Threshold of Hearing



Simultaneous Masking

Fletcher's Model

Consider:

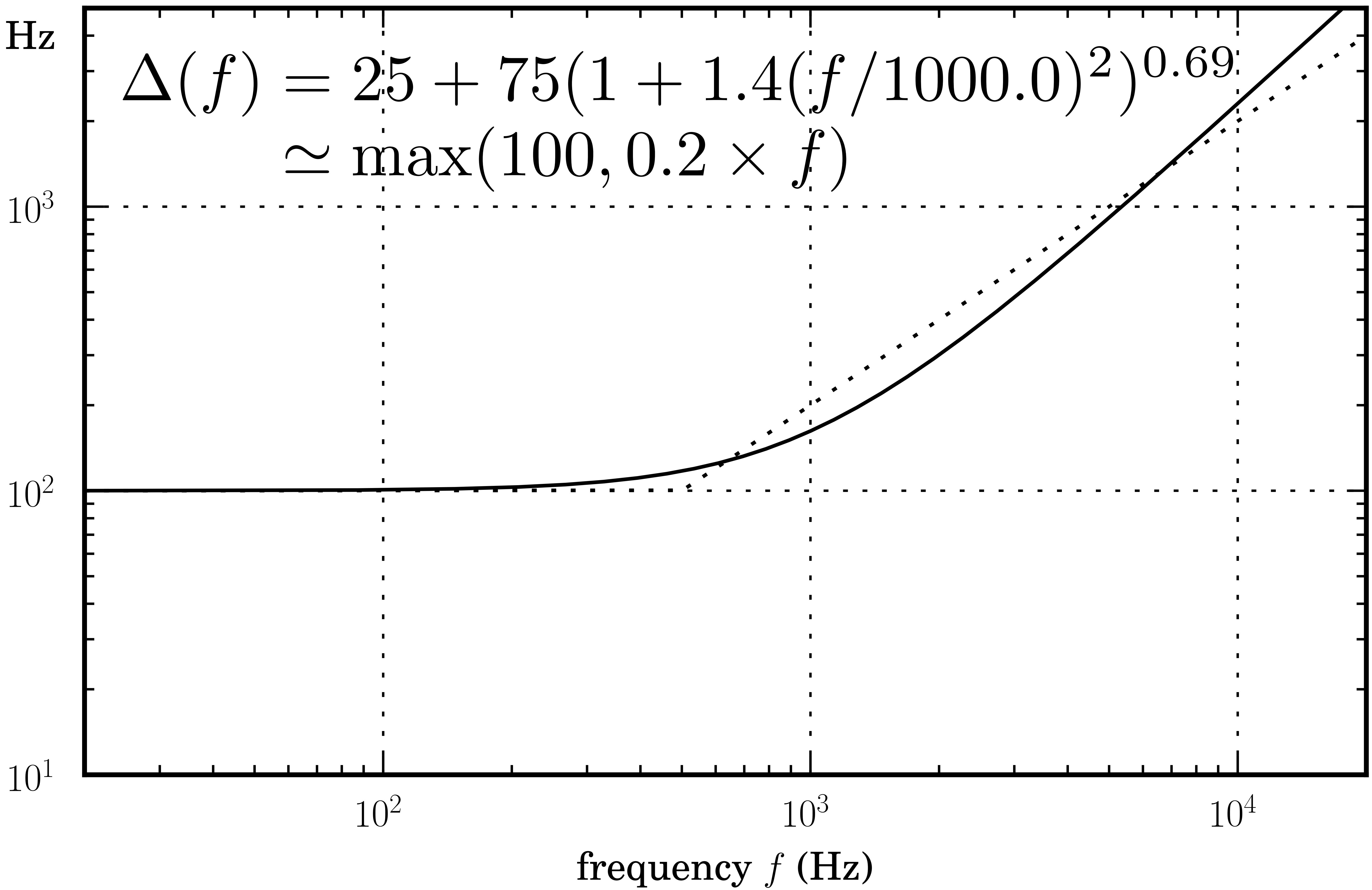
- a masker with sound density $\ell_m(f)$,
- a pure test tone with sound pressure level L and frequency f.

Masking occurs if:

$$\int_{f-\Delta(f)/2}^{f+\Delta(f)/2} 10^{\ell_m(f)/10}\, df \geq 10^{L/10}$$

where $\Delta(f)$ is the **critical bandwidth**.

Critical Bandwidth



The Bark Unit

Measure frequencies in the critical-band scale:

- first, 0 Bark corresponds to 0 Hz,
- then, +1 Bark to +$\Delta(f)$ Hz,

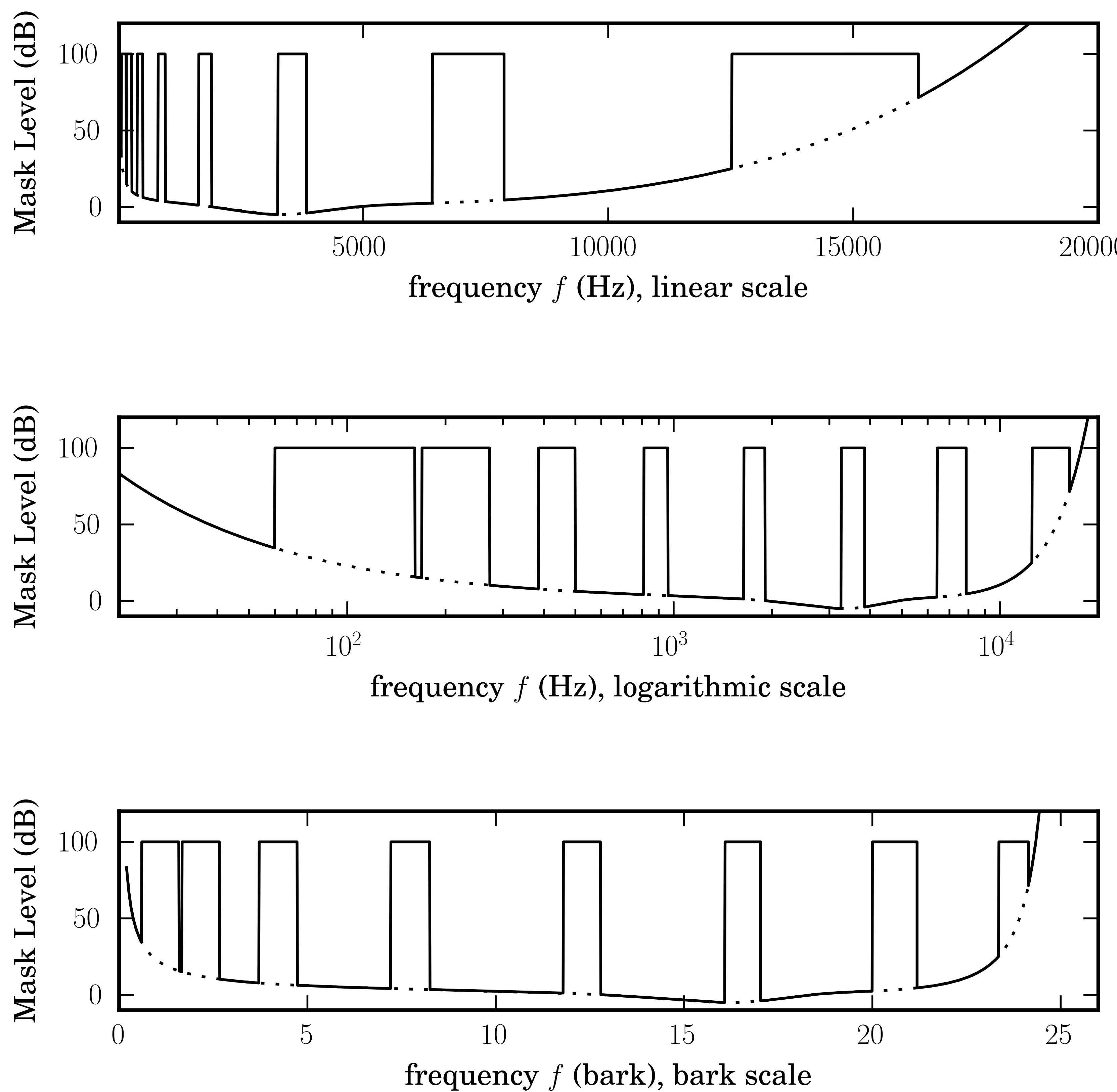
$$
\begin{aligned}
f\,[\mathrm{Bark}] \quad = \quad & 13.0 \times \arctan(0.76 f / 1000.0) \\
& + 3.5 \times \arctan(f / 1000.0 / 7.5)^2
\end{aligned}
$$

```
>>> from psychoacoustics import *
>>> bark([0.0, 1e4, 2e4])
array([0.0,  22.424,  24.575])
>>> hertz([0, 1, 2, 3])
array([0.0,  101.3,  203.7,  308.5])
```

```
>>> critical_bandwidth(440)
113.497
>>> critical_bandwidth(5000)
914.016
```

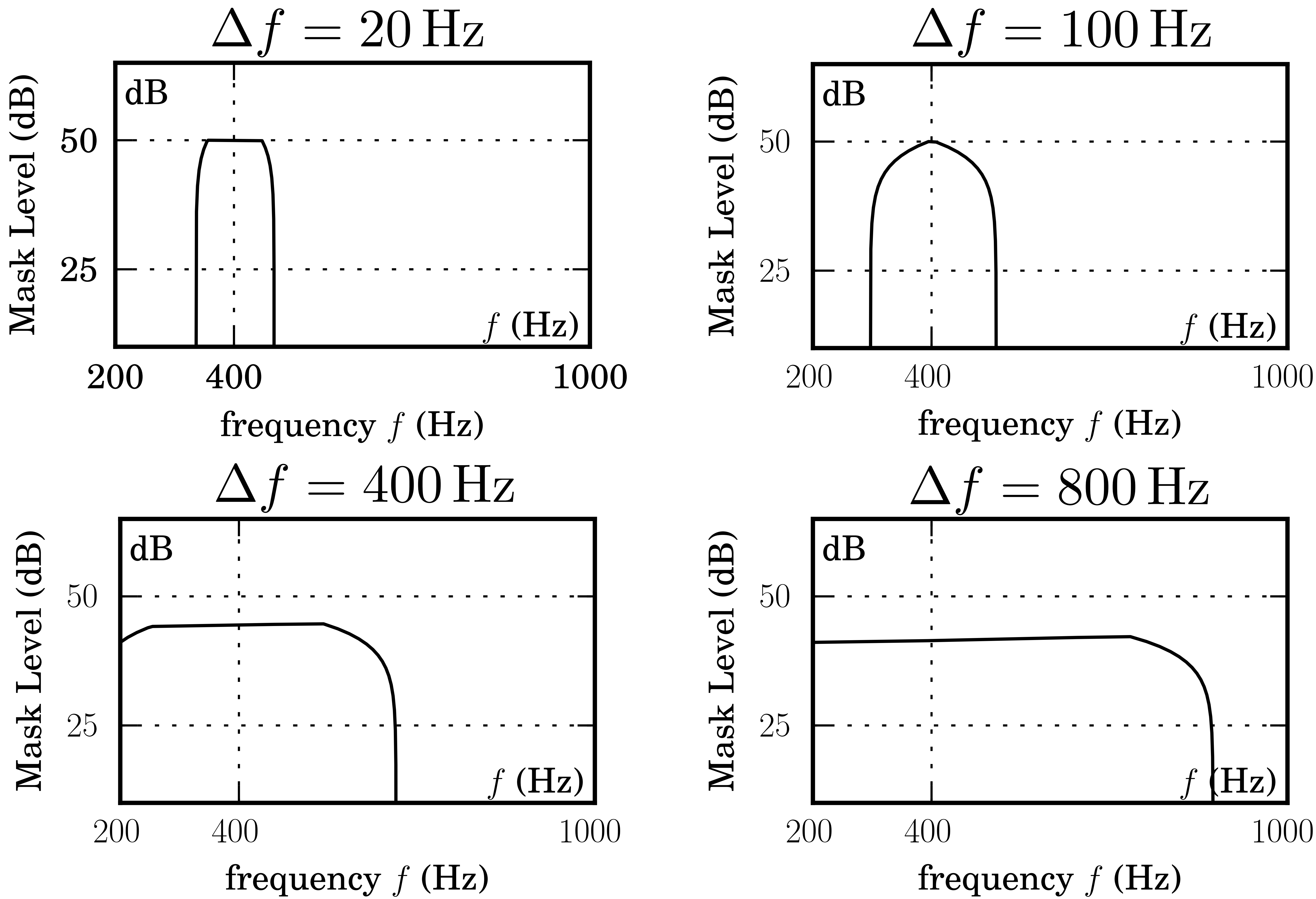
Masking by Pure Tones

Eight pure tones, with the same SPL of $L = 100\,\mathrm{dB}$. The lowest frequency is 110 Hz, and doubles with each new masker.



Band-Limited Noise Maskers

Noise centered at 400 Hz, bandwith Δf, $L = 50\,\text{dB}$.



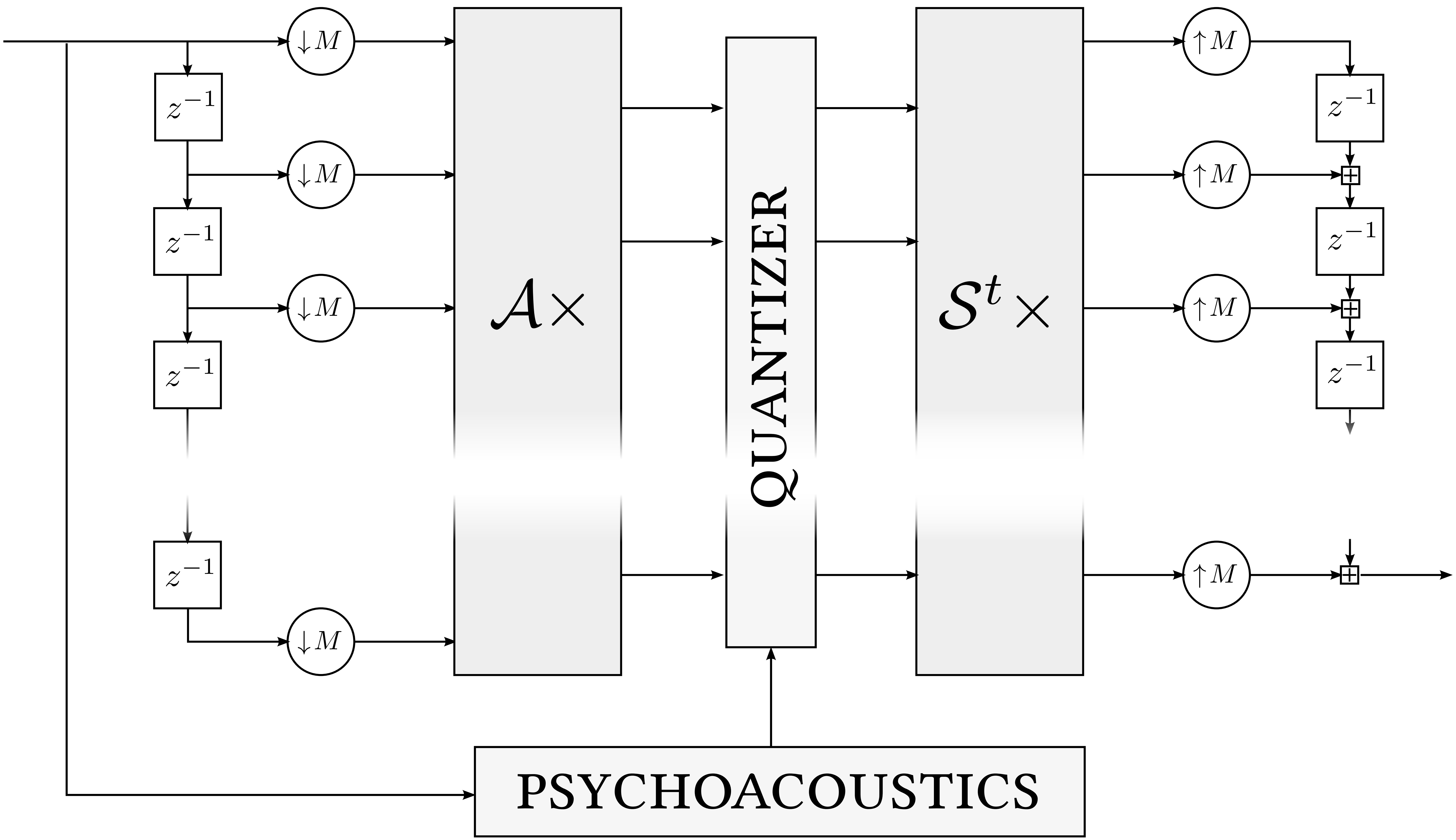
psychoacoustics

Mask class that acts as a compositor.

```
>>> ATH(440.0)
6.9720432781188926
>>> masker = mask(L=50, fc=400, bandwidth=0)
>>> masker(440.0)
50.0
>>> (masker + ATH)(440)
50.00021626
```

sum of intensities

Filter Banks: Bit Allocation



Filter Banks: Bit Allocation

Let X_k be the signal component in subband k, $P_m(k)$ the (normalized) masking level intensity and $[\cdot]_k$ the corresponding quantizer.

The quantization noise is inaudible if:

$$\forall\, k,\ \mathbb{E}[(X_k - [X_k]_k)^2] \leq P_m(k)$$

Meeting these conditions may require high and/or variable bit rates, so we solve instead:

$$\min \sum_{k=0}^{M-1} \frac{\mathbb{E}[(X_k - [X_k]_k)^2]}{P_m(k)}$$

Filter Banks: Bit Allocation

If every $[\cdot]_k$ is a quantizer on $[-1, +1]$ with the same characteristic function f, and b_k bits, the optimal bit allocation is:

$$2^{b_k} \propto \mathrm{SMR}_k \quad \text{with} \quad \mathrm{SMR}_k^2 = \frac{\mathbb{E}[X_k^2]}{P_m(k)}$$

SMR_k is the **signal-to-mask ratio** in subband k.