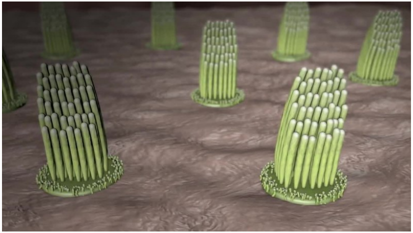
Lecture 3: Loudness, Its Definition, Measurement, and Calculation, Part 2

Mark Hasegawa-Johnson

ECE 537: Speech Processing Fundamentals

Outline

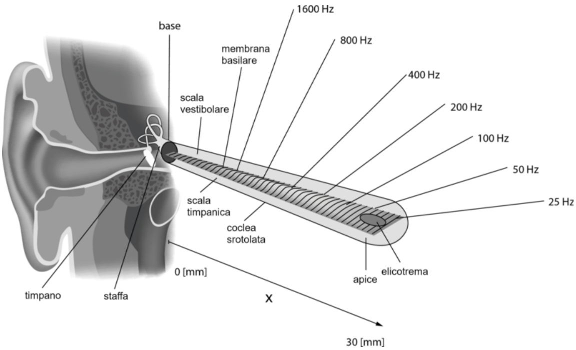
Journey of Sound to the Brain



Public Domain, https://en.wikipedia.org/wiki/File:Journey_of_Sound_to_the_Brain.ogg

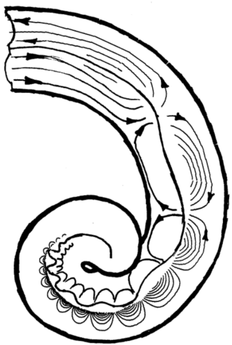
→Watch Video←

Basilar Membrane



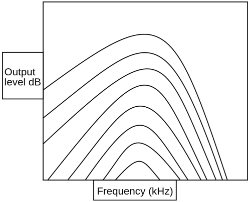
The cochlea is a snail-shaped hole in your skull. The basilar membrane is an epithelium-covered collagen membrane running down the center. The basilar membrane is narrow and taut at the base, wide and floppy at the apex. The resonant frequency of each section is $\omega_{\text{res}} = \sqrt{k/m}$ (k = stiffness, m = mass).

- When a pure tone enters through the oval window, it travels down the basilar membrane until it reaches the spot tuned to its frequency.
- At that spot, all the energy of the pure tone is absorbed by the high-amplitude vibration of the basilar membrane; the tone dies away at that spot, and doesn't travel any farther.



Public Domain, https://en.wikipedia.org/wiki/File:Cochlea_Traveling_Wave.png

Auditory Filters: Symmetric at Low Intensity, Skewed Left at High Intensity



Public Domain,

https://commons.wikimedia.org/wiki/File:OutputlevelMoore.svg

- The auditory filter at frequency f_1 also reacts to loud tones at frequency f_2 if $f_2 < f_1$ (f_2 is more apical than f_1), because the tone at f_2 still has lots of energy when it passes over the part of the basilar membrane tuned to f_1.
- The filter at f_1 doesn't respond as well to frequency f_2 if $f_2 > f_1$ (f_2 is more basal than f_1), because the tone at f_2 has been absorbed by the basilar membrane before it ever reaches the part of the basilar membrane tuned to f_1.

Outline

Adding Intensities

If an acoustic signal is composed of two pure tones, at two different frequencies, their intensities add. Here's the proof, in case you didn't already know it:

$$p(t) = A_1 \cos\left(\frac{2\pi t}{T_1} + \theta_1\right) + A_2 \cos\left(\frac{2\pi t}{T_2} + \theta_2\right)$$

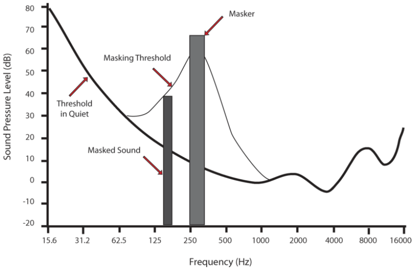
If the least common multiple of T_1 and T_2 is T, we can find the intensity by:

$$\begin{aligned} J &= \frac{1}{\rho c}\frac{1}{T}\int_0^T p^2(t)dt \\ &= \frac{A_1^2}{\rho c} + \frac{A_2^2}{\rho c} = J_1 + J_2 \end{aligned}$$

Masking

Suppose we add a masker with intensity J_m and a signal with intensity J_s. If $J_s + J_m$ differs from J_m by no more than about 1dB, then the signal is **masked** (listener can't distinguish the two signals):

$$10\log_{10}(J_s + J_m) < 10\log_{10}(J_m) + 1\text{dB}$$
$$10\log_{10} J_s < 10\log_{10} J_m - 6\text{dB}$$



Public Domain,

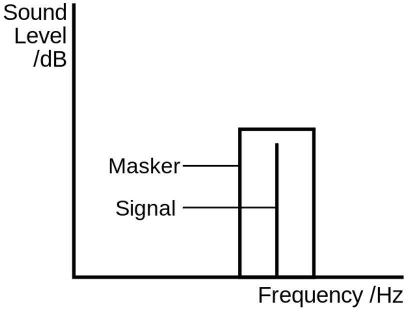
`https://commons.wikimedia.org/wiki/File:Audio_Mask_Graph.png`

Narrowband On-Frequency Masking

Suppose that the masker is bandpass noise, with a bandwidth of B. We can still calculate its intensity, by integrating its power spectrum:

$$J_m = \frac{1}{\rho c} \int E\left[|P(f)|^2\right] df$$

. . . and if $\beta_s < \beta_m - 6\text{dB}$ (roughly), the signal is masked.



Public Domain,

https://commons.wikimedia.org/wiki/File:

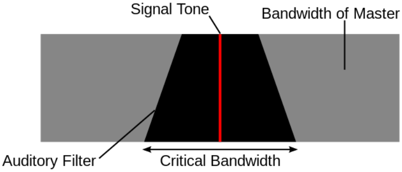
Auditoryfiltermaskersignal1.svg

Wideband On-Frequency Masking

If the masker is wideband noise, the only part of its power that masks the signal is the part that goes through the same auditory filter:

$$J_m = \frac{1}{\rho c} \int |H(f)|^2 E\left[|P(f)|^2\right] df$$

…and if $\beta_s < \beta_m - 6\text{dB}$ (roughly), the signal is masked.



Public Domain,

`https://commons.wikimedia.org/wiki/File:`

`Maskercriticalbandwidth1.svg`

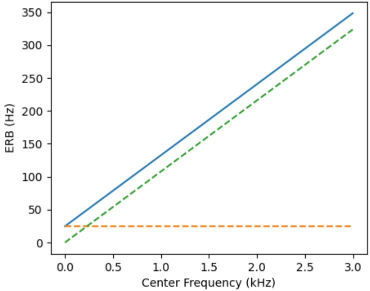
Equivalent Rectangular Bandwidth

The **equivalent rectangular bandwidth** is defined to be the bandwidth of a rectangular filter, $H(f)$, such that masking is predicted by

$$J_m = \frac{1}{\rho c} \int |H(f)|^2 E\left[|P(f)|^2\right] df$$

Moore & Glasberg (1990) used a notched-noise experiment to estimate ERB as a function of center frequency (f in kHz):

$$\text{ERB}(f) = 24.7\,(4.37f + 1)$$



ERB as function of frequency. ERB is 24.7Hz at $f = 0$, and rises linearly from there.

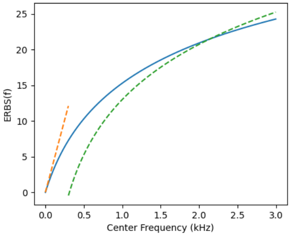
Equivalent Rectangular Bandwidth Scale

The **equivalent rectangular bandwidth scale** is a frequency scale. It counts the number of ERBs that separate the tone f from 0Hz. In other words, $\text{ERBS}(0) = 0$, and for $f > 0$,

$$\frac{df}{d\text{ERBS}(f)} = \text{ERB}(f)$$

Plugging in $\text{ERB}(f) = 24.7(4.37f + 1)$, we find that

$$\text{ERBS}(f) = 11.17 \ln\left(\frac{f + 0.312}{f + 14.675}\right) + 43.0$$

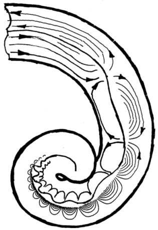


ERBS as function of frequency. ERBS is linear at low frequencies, and logarithmic at high frequencies.

Equivalent Rectangular Bandwidth Scale

$$\text{ERBS}(f) = 11.17 \ln\left(\frac{f+0.312}{f+14.675}\right) + 43.0$$

In a beautiful example of converging scientific evidence, it has been shown that ERBS(f) is approximately equal to the number of millimeters between f and the apex of the basilar membrane. In other words, one ERB is approximately equal to one millimeter on the basilar membrane.



Public Domain, https://en.wikipedia.org/wiki/File:Cochlea_Traveling_Wave.png

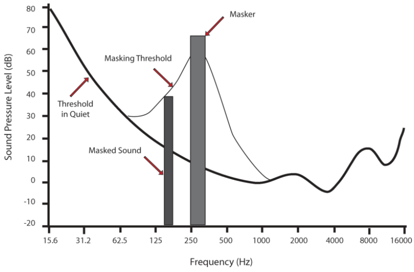
Outline

Partial Masking

Even if the signal is loud enough to be unmasked, its effective loudness is reduced because of the masker. Fletcher & Munson modeled this phenomenon by the equation

$$G = \sum_k b_k G(L_k),$$

where b_2 can be less than one if f_1 and f_2 are within the same critical band.



Public Domain,

https://commons.wikimedia.org/wiki/File:Audio_Mask_Graph.png

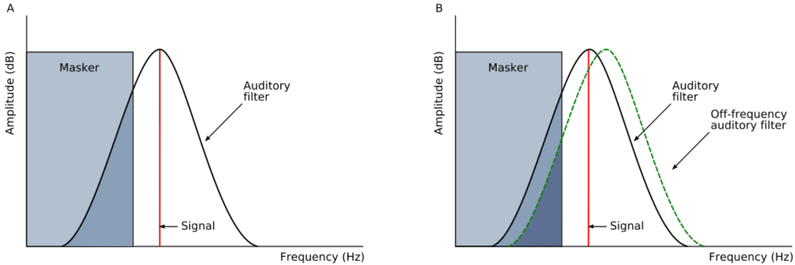
. . . or Adding Loudnesses

How loud is the sum of pure tones? It depends.

- If f_1 and f_2 are **close together** (e.g., $|f_2 - f_1| < 100$Hz), then the loudness is
$$G = G(L(J_1 + J_2)),$$
i.e., add the intensities, find the corresponding loudness level, and convert it to loudness.
- If f_1 and f_2 are **far apart**, however (e.g., $|f_2 - f_1| > 1000$Hz), then the tone is louder:
$$G = G(L(J_1)) + G(L(J_2)) > G(L(J_1 + J_2))$$
- In between those two extremes, the total loudness is $G(L_1) + b_2 G(L_2)$, where b_2 gradually climbs up toward 1.0 as f_2 and f_1 become farther apart.

Off-Frequency Listening

The value of b_2 is determined by the critical band in which the SNR ($= J_s/J_m$) is highest (J_s, J_m = intensity passed by the auditory filter).



The formula for b_k

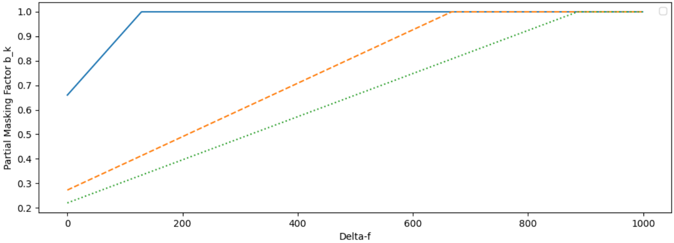
- If $\Delta f = |f_2 - f_1| < B_{\text{center}}$, then just add the intensities of the two tones, and calculate loudness from that. ($B_{\text{center}} \in \{100, 200, 400, 800\}$, depending on f_2).
- If $\Delta f \geq B_{\text{center}}$, then

$$b_2 = \left[\frac{250 + \Delta f}{1000}\right] Q(x)$$

where $Q(x)$ is a function of $x = \beta_1 + 30 \log_{10} f_1 - 95 \approx L_1$.

The formula for b_k

$$b_2 = \left[\frac{250+\Delta f}{1000}\right] Q(x)$$



Solid: $x = 20$dB. Dash: $x = 40$dB. Dot: $x = 60$dB.
$x = \beta_1 + 30\log_{10} f_1 - 95 \approx L$.

Outline

Conclusions

- A critical band is about 1mm on the basilar membrane.
- Within a critical band, intensities add. Therefore, a tone is masked if it doesn't change the level in the critical band by more than 1dB (i.e., signal level is 6dB below masker level).
- Critical bandwidths are roughly $25 + f/10$Hz.

Conclusions

- If two tones are far enough apart,

$$G = \sum_k b_k G(L_k)$$

- The masking factor, b_k, varies from $b_k \geq 0.25$ if $\Delta f = 0$ to $b_k = 1$ if $\Delta f > 1000$:

$$b_2 = \left[\frac{250 + \Delta f}{1000}\right] Q(x)$$

where $Q(x) \approx 2.6$ at 20dB, and $Q(x) \approx 1$ for levels of 40dB or more.