

Lecture 22: Loudness, Its Definition, Measurement, and Calculation

Mark Hasegawa-Johnson

ECE 537: Speech Processing Fundamentals

- ① Fletcher's Model of Hearing
- ② Loudness Level
- ③ Loudness
- ④ Total Masking
- ⑤ Partial Masking
- ⑥ Conclusions

Outline

- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness
- 4 Total Masking
- 5 Partial Masking
- 6 Conclusions

Loudness adds!

The most revolutionary new idea in Fletcher & Munson's paper was the idea that the loudnesses of different tones add together. If there's a tone at frequency f_1 with loudness level L_1 , and a tone at frequency f_2 with loudness level L_2 , and if f_2 and f_1 are far enough apart, then

$$G(L) = G(L_1) + G(L_2)$$

where $G(L)$ is the perceived loudness of the whole signal, $G(L_1)$ is the loudness of the first tone, and $G(L_2)$ is the loudness of the second tone.

Loudness adds, except when it doesn't

Fletcher's model of hearing was based on experiments showing that the loudness of a composite tone is equal to the sum of the loudnesses of the individual component tones, unless the frequencies are close together. If they're close together, then the total loudness is less than the sum of the component loudnesses:

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

where $b_k \leq 1$. If the frequencies are far apart, then $b_k = 1$. Friday's lecture will consider what happens when they are close together.

Outline

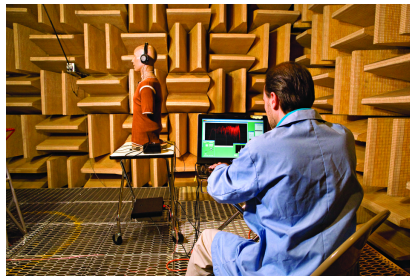
- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness
- 4 Total Masking
- 5 Partial Masking
- 6 Conclusions

Listening Tests

So how do you measure the “loudness” of a sound?

First idea: let’s find out how it depends on frequency.

Play two different tones to a person, and ask them to adjust the level of the second tone until it sounds exactly as loud as the first tone.



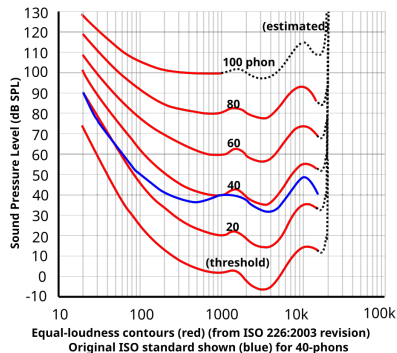
CC-SA 4.0,

[https://commons.wikimedia.org/wiki/File:
Consumer_Reports_-_product_testing_-_
headphones_in_anechoic_chamber.tif](https://commons.wikimedia.org/wiki/File:Consumer_Reports_-_product_testing_-_headphones_in_anechoic_chamber.tif)

Loudness Level

The “loudness level” of any pure tone is defined to be the level, in decibels, of a 1000Hz tone that sounds equally loud.

The figure showing loudness levels, as a function of level and frequency, is perhaps the most famous figure from Fletcher & Munson's paper. It's now an ISO standard.



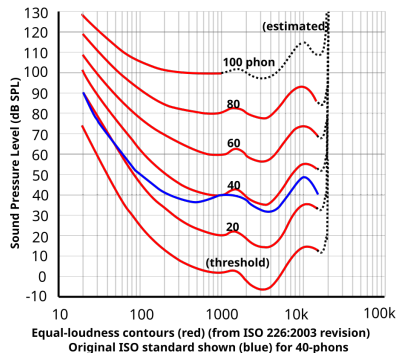
Public domain, <https://commons.wikimedia.org/wiki/File:Lindos1.svg>

[//commons.wikimedia.org/wiki/File:Lindos1.svg](https://commons.wikimedia.org/wiki/File:Lindos1.svg)

Loudness Level

So if we have a sound with some level β (the y-axis in the figure) at frequency f (the x-axis in the figure), and if the equally loud 1000Hz tone has a level $\beta_1(f, \beta)$ (the numbers written on the curves in the figure), then the **loudness level** (measured in **phons**) is

$$L = \beta_1(f, \beta)$$



Outline

- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness**
- 4 Total Masking
- 5 Partial Masking
- 6 Conclusions

From “Loudness Level” to “Loudness”

Now we know how to normalize for frequency: we can find the loudness level $L(\beta, f)$.

But Fletcher & Munson proposed a harder problem: they proposed to find a transformation $G(L)$ such that

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

Let's call $G(L)$ the **loudness** (measured in **sones**).
How can we find $G(L)$?

How: Assume that loudness adds!

Suppose that we have two tones with exactly the same loudness level, $L_1 = L_2$. If the model is true — if loudnesses really add — then the loudness of the composite tone should be just exactly twice the loudness of either composite tone!

$$G(L) = G(L_1) + G(L_2) = 2G(L_1)$$

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 J (measured in Watts/ m^2) as

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

Adding intensities < Adding loudnesses

- So the intensity of the composite tone is the sum of the intensities of the two components: $J = J_1 + J_2$.
- But here's a problem: if you set a 1000Hz pure-tone to have the same intensity, J , it would **sound quieter!**
- If you adjust the intensity of the 1000Hz pure-tone so it sounds equally loud as the composite tone, you would have to adjust it to an intensity louder than J .

How: Use the fact that loudness adds

So the experiment is as follows:

- Play a composite tone, composed of two components with the same loudness level $L_1 = L_2$.
- Ask the listener to adjust the level of a pure 1000Hz tone until it sounds exactly as loud as the composite tone.
- Let the level of the 1000Hz tone be L . Now we know that

$$G(L) = 2G(L_1)$$

Measurement results

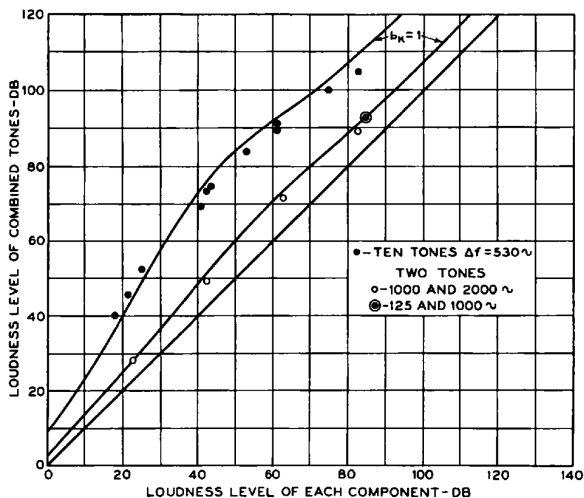
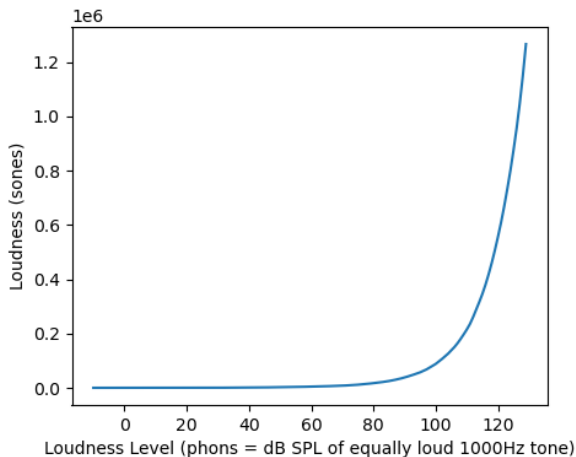


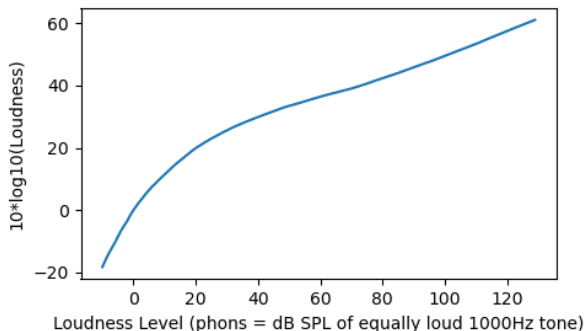
Fig. 6—Complex tones having components widely separated in frequency.

Loudness as a function of Loudness Level



This is the data from Table III in the article (loudness as a function of loudness level).

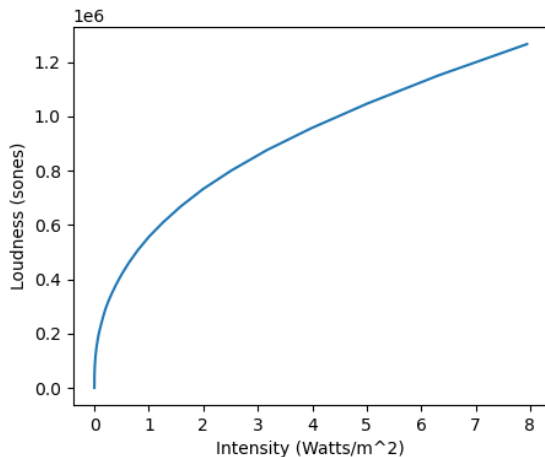
Log-Loudness as a function of Loudness Level



Is Loudness = Intensity?

- On the previous slide, notice that, above about 30phons, the logarithm of loudness is proportional to loudness level.
- Remember that loudness level is the logarithm of intensity (of an equivalent 1000Hz tone).
- Does that mean that loudness = intensity?

Loudness as a function of Intensity (of an equivalently loud 1000Hz tone)



Loudness is a Compressive Function of Intensity

- Loudness is a compressive function of intensity. . .
- . . . but it's not quite as compressive as a logarithm.
- A good model is that $G \propto J^{1/3}$, where the constant of proportionality depends on frequency ($G \propto 10^{L/30}$)

Loudness is a Compressive Function of Intensity

If you adjust the intensity of a 1000Hz pure tone to be J , where $J = J_1 + J_2$ is the same as the total intensity of a composite tone, you will find that the **pure tone is quieter**.

That's because loudness is a compressive nonlinearity:

$$G(J_1 + J_2) < G(J_1) + G(J_2)$$

For example, the cube-root function has this property.

Outline

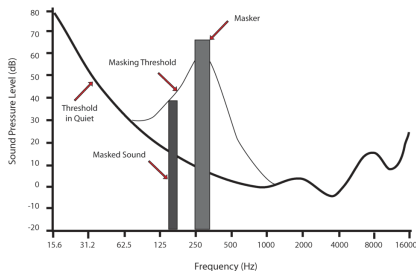
- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness
- 4 Total Masking**
- 5 Partial Masking
- 6 Conclusions

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:](https://commons.wikimedia.org/wiki/File:Audio_Mask_Graph.png)

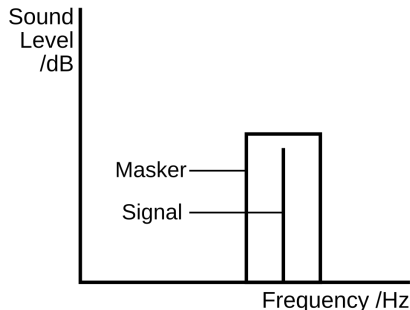
Audio_Mask_Graph.png

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:

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

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



Public Domain,

[https://commons.wikimedia.org/wiki/File:](https://commons.wikimedia.org/wiki/File:Auditoryfiltermaskersignal1.svg)

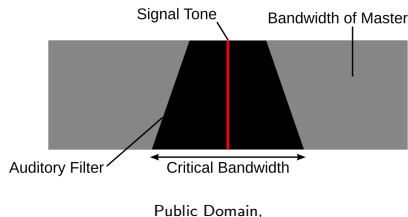
Auditoryfiltermaskersignal1.svg

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:

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

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



[https://commons.wikimedia.org/wiki/File:
Maskercriticalbandwidth1.svg](https://commons.wikimedia.org/wiki/File:Maskercriticalbandwidth1.svg)

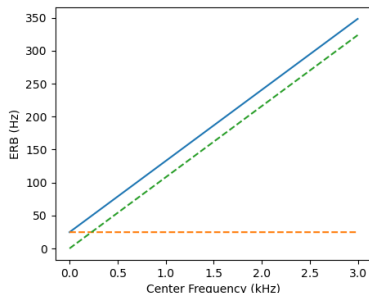
Equivalent Rectangular Bandwidth

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

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

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

$$\text{ERB}(f) = 24.7 \left(4.37 \left(\frac{f}{1000} \right) + 1 \right)$$



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

Equivalent Rectangular Bandwidth Scale

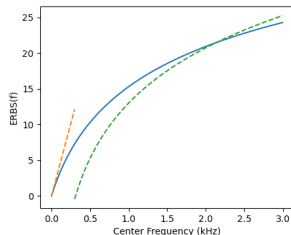
The **equivalent rectangular bandwidth scale** is a frequency scale. It counts the number of ERBs that separate the tone f from 0Hz. In other words, $ERBS(0) = 0$, and for $f > 0$,

$$\frac{df}{dERBS(f)} = ERB(f)$$

Plugging in

$ERB(f) = 24.7 \left(4.37 \left(\frac{f}{1000} \right) + 1 \right)$, we find that

$$ERBS(f) = 11.17 \ln \left(\frac{f + 312}{f + 14675} \right) + 43.0$$



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

Equivalent Rectangular Bandwidth Scale

$$\text{ERBS}(f) = 11.17 \ln \left(\frac{f + 312}{f + 14675} \right) + 43.0$$

In a beautiful example of converging scientific evidence, it has been shown that $\text{ERBS}(f)$ 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.



Outline

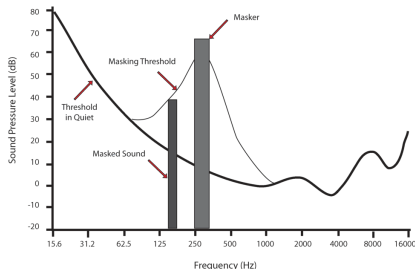
- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness
- 4 Total Masking
- 5 Partial Masking**
- 6 Conclusions

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.



...or Adding Loudnesses

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

- If f_1 and f_2 are **close together** (e.g., $|f_2 - f_1| < 100\text{Hz}$), 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.

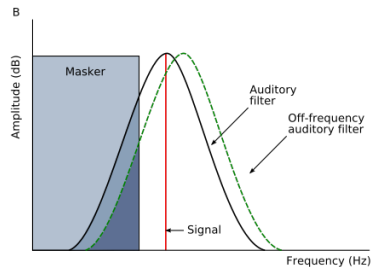
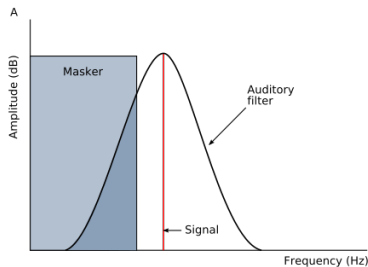
- If f_1 and f_2 are **far apart**, however (e.g., $|f_2 - f_1| > 1000\text{Hz}$), then the tone is louder:

$$G = G(L(J_1)) + G(L(J_2)) > G(L(J_1 + J_2))$$

- 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

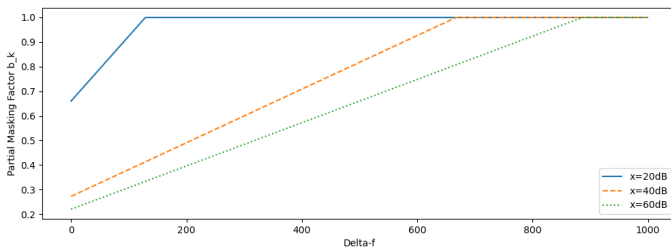
The value of b_2 is determined by the critical band in which the SNR ($= J_s/J_m$) is highest (J_s, J_m = intensity passed by the auditory filter).



Public Domain, https://commons.wikimedia.org/wiki/File:Off_F_listening.svg

Fletcher's formula for b_2

$$b_2 = \min \left(1, \frac{250 + |f_2 - f_1|}{1000} Q(L_1) \right)$$



... where $Q(L_1)$ is a slope that depends on the level of the masker.
Solid: $L_1 = 20\text{dB}$. Dash: $L_1 = 40\text{dB}$. Dot: $L_1 = 60\text{dB}$.

Outline

- 1 Fletcher's Model of Hearing
- 2 Loudness Level
- 3 Loudness
- 4 Total Masking
- 5 Partial Masking
- 6 Conclusions**

Conclusions (1/3)

- Loudness Level (phons) is defined to be the level (in dB SPL) of an equally loud 1000Hz tone.
- Loudness (sones) is defined by the equation

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

... where $b_k \leq 1$, and for tones widely spaced in frequency, $b_k = 1$. A not-too-bad approximate model (for tones louder than about 30dB SPL) is $G \propto |J|^{1/3}$.

Conclusions (2/3)

- 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 $25 + f/10\text{Hz}$.
- The number of millimeters a tone is away from the apex of the basilar membrane is roughly

$$\text{ERBS}(f) = 11.17 \ln \left(\frac{f + 312}{f + 14675} \right) + 43.0$$

Conclusions (3/3)

- If two tones are far enough apart,

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

- The masking factor, b_2 , varies from about $0.25 \leq b_2 \leq 1$, roughly as:

$$b_2 = \min \left(1, \frac{250 + |f_2 - f_1|}{1000} Q(L_1) \right)$$

where $Q(L_1) \approx 2.6$ at $L_1 = 20\text{phons}$, and $Q(L_1) \approx 1$ for $L_1 \geq 40\text{phons}$.