Lecture 4: Filtered Noise

Mark Hasegawa-Johnson

ECE 417: Multimedia Signal Processing, Fall 2020

Outline

Review: Last time

- Masking: a pure tone can be heard, in noise, if there is **at least one** auditory filter through which $\frac{N_k+T_k}{N_k} >$ threshold.
- We can calculate the power of a noise signal by using Parseval's theorem, together with its power spectrum.

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = \frac{1}{N}\sum_{k=0}^{N-1} R[k] = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\omega)d\omega$$

- The inverse DTFT of the power spectrum is the autocorrelation

$$r[n] = \frac{1}{N}x[n] * x[-n]$$

- The power spectrum and autocorrelation of noise are, themselves, random variables. For zero-mean white noise of length N, their expected values are

$$E\left[R[k]\right] = \sigma^2$$

$$E\left[r[n]\right] = \sigma^2\delta[n]$$

Outline

Filtered Noise

What happens when we filter noise? Suppose that $x[n]$ is zero-mean Gaussian white noise, and

$$y[n] = h[n] * x[n]$$

What is $y[n]$?

Filtered Noise

$$y[n] = h[n] * x[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m]$$

- $y[n]$ is the sum of Gaussians, so $y[n]$ is also Gaussian.
- $y[n]$ is the sum of zero-mean random variables, so it's also zero-mean.
- $y[n] = h[0]x[n] +$ other stuff, and $y[n+1] = h[1]x[n] +$ other stuff. So obviously, $y[n]$ and $y[n+1]$ are not uncorrelated. So $y[n]$ is not white noise.
- What kind of noise is it?

The variance of $y[n]$

First, let's find its variance. Since $x[n]$ and $x[n+1]$ are uncorrelated, we can write

$$\sigma_y^2 = \sum_{m=-\infty}^{\infty} h^2[m]\mathrm{Var}(x[n-m])$$

$$= \sigma_x^2 \sum_{m=-\infty}^{\infty} h^2[m]$$

The autocorrelation of $y[n]$

Second, let's find its autocorrelation. Let's define $r_{xx}[n] = \frac{1}{N} x[n] * x[-n]$. Then

$$
\begin{aligned}
r_{yy}[n] &= \frac{1}{N} y[n] * y[-n] \\
&= \frac{1}{N} (x[n] * h[n]) * (x[-n] * h[-n]) \\
&= \frac{1}{N} x[n] * x[-n] * h[n] * h[-n] \\
&= r_{xx}[n] * h[n] * h[-n]
\end{aligned}
$$

Expected autocorrelation of $y[n]$

$$r_{yy}[n] = r_{xx}[n] * h[n] * h[-n]$$

Expectation is linear, and convolution is linear, so

$$E\left[r_{yy}[n]\right] = E\left[r_{xx}[n]\right] * h[n] * h[-n]$$

Expected autocorrelation of $y[n]$

$x[n]$ is white noise if and only if its autocorrelation is a delta function:

$$E\left[r_{xx}[n]\right] = \sigma_x^2 \delta[n]$$

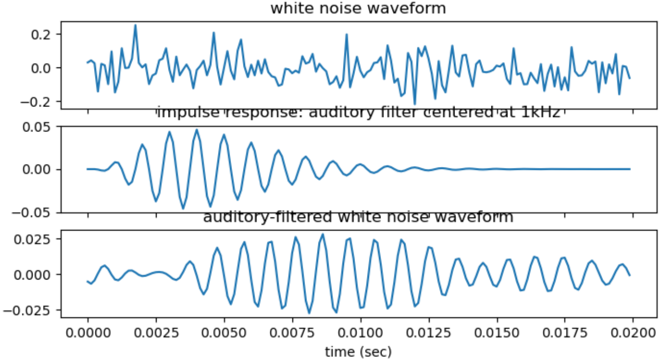
So

$$E\left[r_{yy}[n]\right] = \sigma_x^2 \left(h[n] * h[-n]\right)$$

In other words, $x[n]$ contributes only its energy (σ^2). $h[n]$ contributes the correlation between neighboring samples.

Example

Here's an example. The white noise signal on the top ($x[n]$) is convolved with the bandpass filter in the middle ($h[n]$) to produce the green-noise signal on the bottom ($y[n]$). Notice that $y[n]$ is random, but correlated.



Colors, anybody?

- Noise with a flat power spectrum (uncorrelated samples) is called white noise.
- Noise that has been filtered (correlated samples) is called colored noise.
    - If it's a low-pass filter, we call it pink noise (this is quite standard).
    - If it's a high-pass filter, we could call it blue noise (not so standard).
    - If it's a band-pass filter, we could call it green noise (not at all standard, but I like it!)

Outline

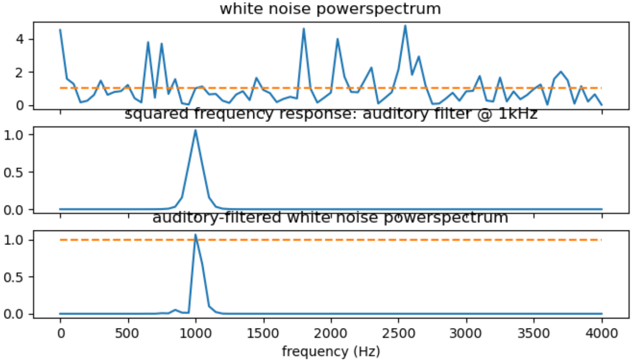
Power Spectrum of Filtered Noise

So we have $r_{yy}[n] = r_{xx}[n] * h[n] * h[-n]$. What about the power spectrum?

$$\begin{aligned} R_{yy}(\omega) &= \mathcal{F}\left\{r_{yy}[n]\right\} \\ &= \mathcal{F}\left\{r_{xx}[n] * h[n] * h[-n]\right\} \\ &= R_{xx}(\omega)|H(\omega)|^2 \end{aligned}$$

Example

Here's an example. The white noise signal on the top ($|X[k]|^2$) is multiplied by the bandpass filter in the middle ($|H[k]|^2$) to produce the green-noise signal on the bottom ($|Y[k]|^2 = |X[k]|^2|H[k]|^2$).



Units Conversion

The DTFT version of Parseval's theorem is

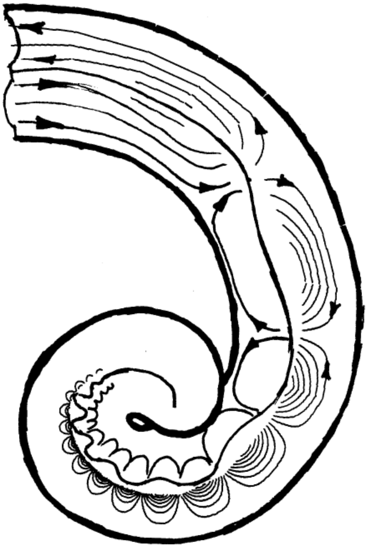
$$\frac{1}{N}\sum_n x^2[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_{xx}(\omega)d\omega$$

Let's consider converting units to Hertz. Remember that $\omega = \frac{2\pi f}{F_s}$, where F_s is the sampling frequency, so $d\omega = \frac{2\pi}{F_s}df$, and we get that

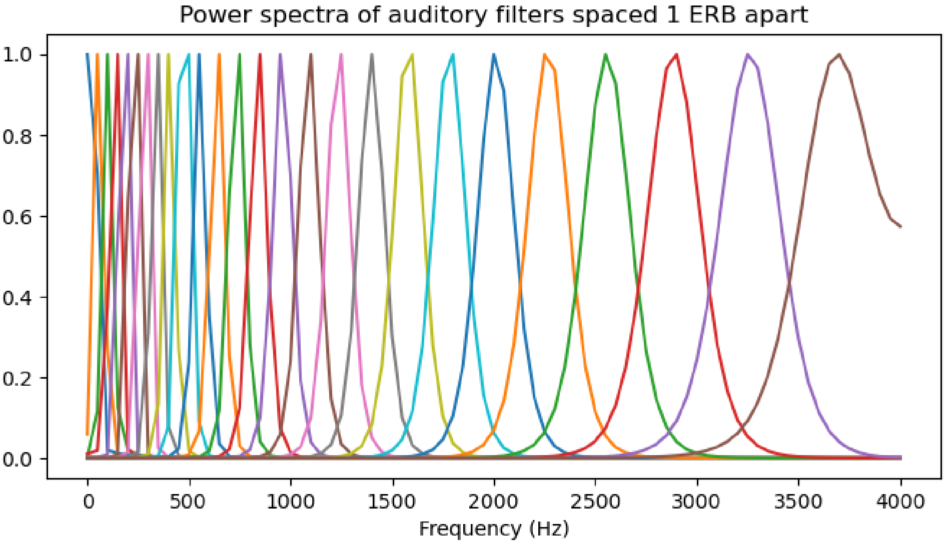
$$\frac{1}{N}\sum_n x^2[n] = \frac{1}{F_s}\int_{-F_s/2}^{F_s/2} R_{xx}\left(\frac{2\pi f}{F_s}\right) df$$

So we can use $R_{xx}\left(\frac{2\pi f}{F_s}\right)$ as if it were a power spectrum in continuous time, at least for $-\frac{F_s}{2} < f < \frac{F_s}{2}$.

Outline



Dick Lyon, public domain image, 2007. https://en.wikipedia.org/wiki/File:Cochlea_Traveling_Wave.png


Power spectra of auditory filters spaced 1 ERB apart
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
3500
4000
Frequency (Hz)

The Power of Filtered White Noise

Suppose that $h[n]$ is the auditory filter centered at frequency f_c (in Hertz), and

$$y[n] = h[n] * x[n]$$

where $x[n]$ is white noise. What's the power of the signal $y[n]$?

$$\begin{aligned}\frac{1}{N}\sum_n y^2[n] &= \frac{1}{F_s}\int_{-F_s/2}^{F_s/2} R_{yy}\left(\frac{2\pi f}{F_s}\right) df \\ &= \frac{1}{F_s}\int_{-F_s/2}^{F_s/2} R_{xx}\left(\frac{2\pi f}{F_s}\right) |H(f)|^2 df\end{aligned}$$

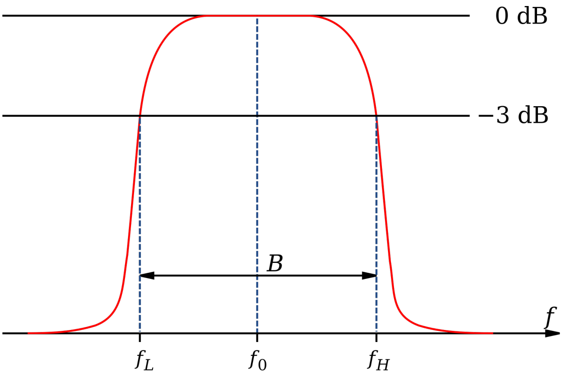
So the expected power is

$$E\left[\frac{1}{N}\sum_n y^2[n]\right] = \frac{\sigma^2}{F_s}\int_{-F_s/2}^{F_s/2} |H(f)|^2 df$$

. . . so, OK, what is $\int_{-F_s/2}^{F_s/2} |H(f)|^2 df$?

Outline

Bandwidth



By InductiveLoad, public domain image, `https://commons.wikimedia.org/wiki/File:Bandwidth_2.svg`

Equivalent rectangular bandwidth

Let's make the simplest possible assumption: a rectangular filter, centered at frequency f_c, with bandwidth b:

$$H(f) = \begin{cases} 1 & f_c - \frac{b}{2} < f < f_c + \frac{b}{2} \\ 1 & f_c - \frac{b}{2} < -f < f_c + \frac{b}{2} \\ 0 & \text{otherwise} \end{cases}$$

That's useful, because it makes Parseval's theorem very easy:

$$\frac{\sigma^2}{F_s} \int_{-F_s/2}^{F_s/2} |H(f)|^2 df = \left(\frac{2b}{F_s}\right) \sigma^2$$

Reminder: Fletcher's Model of Masking

Fletcher proposed the following model of hearing in noise:

1. The human ear pre-processes the audio using a bank of bandpass filters.
2. The power of the noise signal, in the bandpass filter centered at frequency f_c, is N_{f_c}.
3. The power of the noise+tone is $N_{f_c} + T_{f_c}$.
4. If there is **any** band, k, in which $\frac{N_{f_c}+T_{f_c}}{N_{f_c}} >$ threshold, then the tone is audible. Otherwise, not.

The "Just Noticeable Difference" in Loudness

First, let's figure out what the threshold is. Play two white noise signals, $x[n]$ and $y[n]$. Ask listeners which one is louder.

The "just noticeable difference" is the difference in loudness at which 75% of listeners can correctly tell you that $y[n]$ is louder than $x[n]$:

$$\text{JND} = 10\log_{10}\left(\sum_n y^2[n]\right) - 10\log_{10}\left(\sum_n x^2[n]\right)$$

It turns out that the JND is very close to 1dB, for casual listening, for most listeners. So Fletcher's masking criterion becomes:

- If there is **any** band, l, in which $10\log_{10}\left(\frac{N_{f_c}+T_{f_c}}{N_{f_c}}\right) > 1\text{dB}$, then the tone is audible. Otherwise, not.

Fletcher's Model, for White Noise

1. The human ear pre-processes the audio using a bank of bandpass filters.
2. The power of the noise signal, in the filter centered at f_c, is $N_{f_c} = 2b\sigma^2/F_s$.
3. The power of the noise+tone is $N_{f_c} + T_{f_c}$.
4. If there is any band in which $10\log_{10}\left(\frac{N_{f_c}+T_{f_c}}{N_{f_c}}\right) > 1\text{dB}$, then the tone is audible. Otherwise, not.

...next question to solve. What is the power of the tone?

What is the power of a tone?

A pure tone has the formula

$$x[n] = A\cos\left(\omega_0 n + \theta\right), \quad \omega_0 = \frac{2\pi}{N_0}$$

Its power is calculated by averaging over any integer number of periods:

$$T_{f_c} = \frac{1}{N_0}\sum_{n=0}^{N_0-1} A^2\cos^2\left(\omega_0 n + \theta\right) = \frac{A^2}{2}$$

Power of a filtered tone

Suppose $y[n] = h[n] * x[n]$. Then

$$y[n] = A|H(\omega_0)|\cos\left(\omega_0 n + \theta + \angle H_{f_c}(\omega_0)\right)$$

And it has the power

$$T_{f_c} = \frac{1}{2}A^2|H(\omega_0)|^2$$

If we're using rectangular bandpass filters, then

$$T_{f_c} = \begin{cases} \frac{A^2}{2} & f_c - \frac{b}{2} < f_0 < f_c + \frac{b}{2} \\ \frac{A^2}{2} & f_c - \frac{b}{2} < -f < f_c + \frac{b}{2} \\ 0 & \text{otherwise} \end{cases}$$

Fletcher's Model, for White Noise

The tone is audible if there's some filter centered at $f_c \approx f_0$ (specifically, $f_c - \frac{b}{2} < f_0 < f_c + \frac{b}{2}$) for which:

$$1\text{dB} < 10\log_{10}\left(\frac{N_{f_c} + T_{f_c}}{N_{f_c}}\right) = 10\log_{10}\left(\frac{\frac{2b\sigma^2}{F_s} + \frac{A^2}{2}}{\frac{2b\sigma^2}{F_s}}\right)$$

Procedure: Set F_s and σ^2 to some comfortable listening level. In order to find the bandwidth, b, of the auditory filter centered at f_0,

1. Test a range of different levels of A.
2. Find the minimum value of A at which listeners can report "tone is present" with 75% accuracy.
3. From that, calculate b.

Equivalent rectangular bandwidth (ERB)

Here are the experimental results: The frequency resolution of your ear is better at low frequencies! In fact, the dependence is roughly linear (Glasberg and Moore, 1990):

$$b \approx 0.108f + 24.7$$

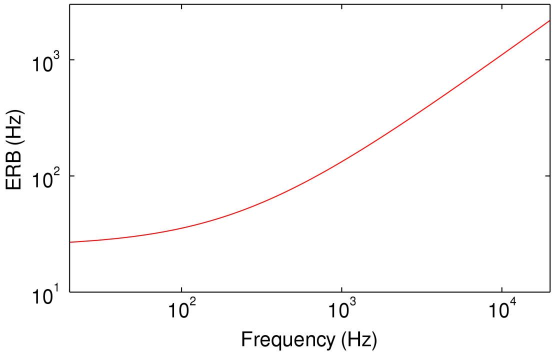
These are often called (approximately) constant-Q filters, because the quality factor is

$$Q = \frac{f}{b} \approx 9.26$$

The dependence of b on f is not quite linear. A more precise formula is given in (Moore and Glasberg, 1983) as:

$$b = 6.23\left(\frac{f}{1000}\right)^2 + 93.39\left(\frac{f}{1000}\right) + 28.52$$

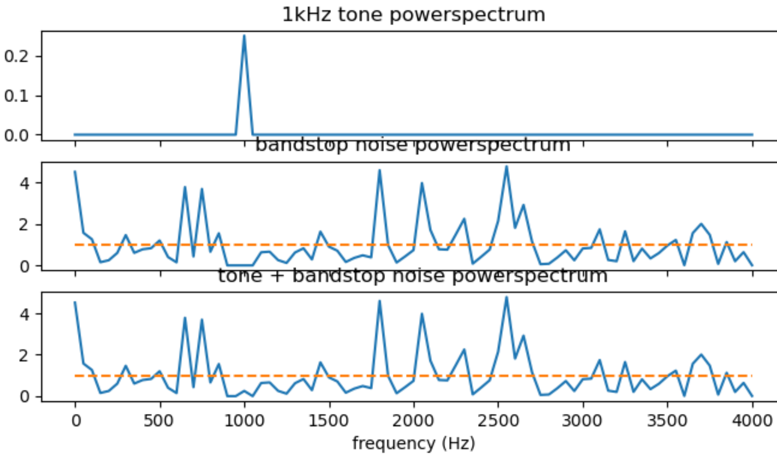
Equivalent rectangular bandwidth (ERB)



By Dick Lyon, public domain image 2009, `https://commons.wikimedia.org/wiki/File:ERB_vs_frequency.svg`

Outline

What happens if we start with bandstop noise?



The power of bandstop noise

Suppose $y[n]$ is a bandstop noise: say, it's been zeroed out between f_2 and f_3:

$$E\left[R_{yy}(\omega)\right] = \begin{cases} 0 & f_2 < |f| < f_3 \\ \sigma^2 & \text{otherwise} \end{cases}$$

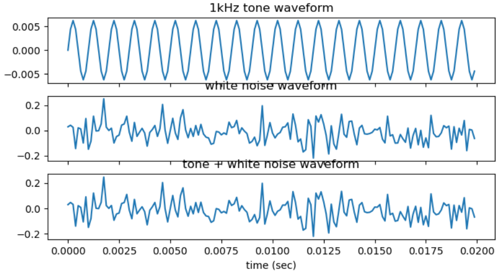
Parseval's theorem gives us the energy of this noise:

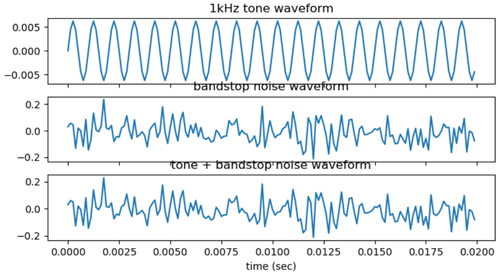
$$E\left[\frac{1}{N}\sum_{n=0}^{N-1} y^2[n]\right] = \frac{1}{F_s}\int_{-F_s/2}^{F_s/2} R_{yy}(\omega)d\omega$$

$$\sigma^2\left(1 - \frac{2(f_3 - f_2)}{F_s}\right)$$

If $f_3 - f_2 \ll \frac{F_s}{2}$, then the power of this noise is almost as large as the power of a white noise signal.

Bandstop noise power ≈ White noise power





Auditory-filtered bandstop noise

Now let's filter $y[n]$ through an auditory filter:

$$z[n] = y[n] * h[n]$$

where, again, let's assume a rectangular auditory filter, and let's assume that the whole bandstop region lies inside the auditory filter, so that

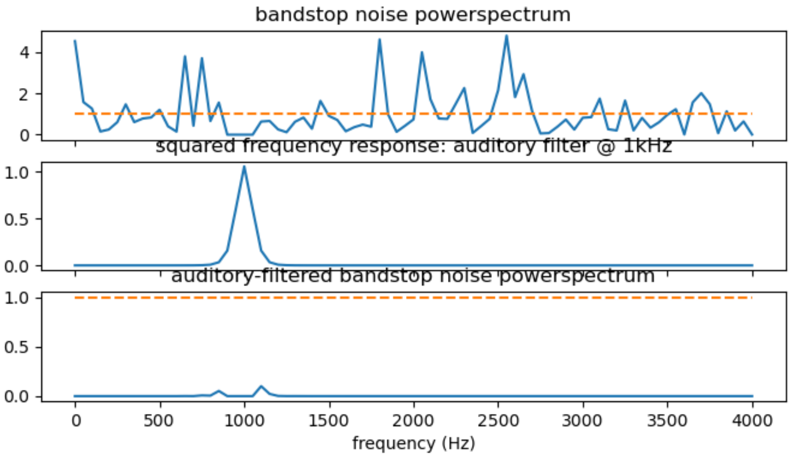
$$f_c - \frac{b}{2} < f_3 < f_2 < f_c + \frac{b}{2}$$

Then we have

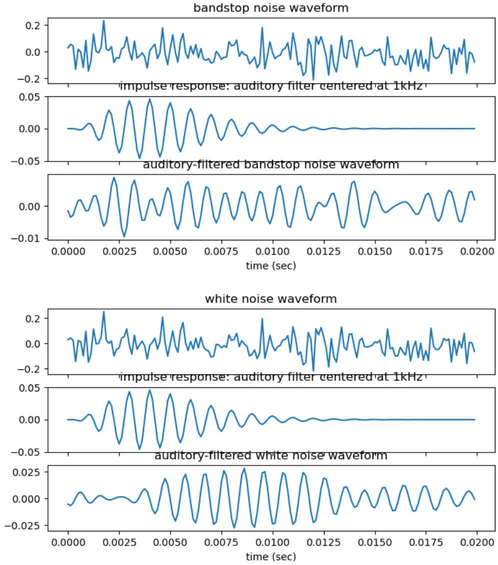
$$E\left[R_{zz}(f)\right] = E\left[R_{yy}(f)\right] |H(f)|^2 = \begin{cases} \sigma^2 & f_c - \frac{b}{2} < |f| < f_2 \\ \sigma^2 & f_3 < |f| < f_c + \frac{b}{2} \\ 0 & \text{otherwise} \end{cases}$$

This is nonzero only in two tiny frequency bands: $f_c - \frac{b}{2} < |f| < f_2$, and $f_3 < |f| < f_c + \frac{b}{2}$.

Auditory-filtered bandstop noise



Tiny power spectrum ⇒ tiny waveform energy



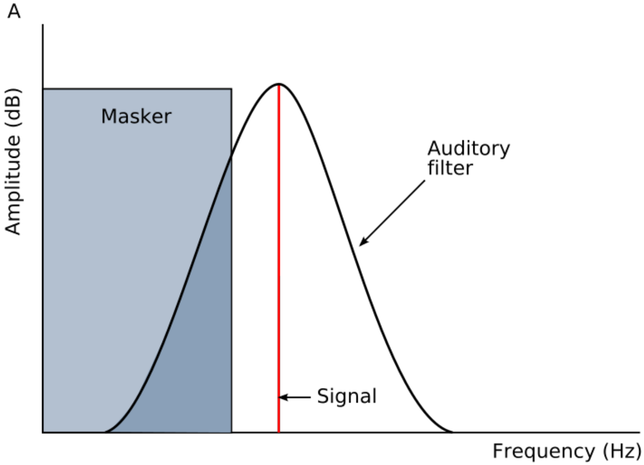
Outline

Lowpass noise

Patterson (1974) measured the shape of the auditory filter using lowpass noise, i.e., noise with the following spectrum:

$$R_{xx}(f) = \begin{cases} \sigma^2 & -f_1 < f < f_1 \\ 0 & \text{otherwise} \end{cases}$$

Lowpass-filtered noise



By Dave Dunford, public domain image 1010, `https://en.wikipedia.org/wiki/File:Off_F_listening.svg`

Lowpass noise

Patterson (1974) measured the shape of the auditory filter using lowpass noise, i.e., noise with the following spectrum:

$$R_{xx}(f) = \begin{cases} \sigma^2 & -f_1 < f < f_1 \\ 0 & \text{otherwise} \end{cases}$$

The power of a lowpass filtered noise, as heard through an auditory filter $H(f)$ centered at f_c, is

$$N(f_c, f_1) = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} R_{xx}(f) |H(f)|^2 df = \frac{\sigma^2}{F_s} \int_{-f_1}^{f_1} |H(f)|^2 df$$

Turning that around, we get a formula for $|H(f)|^2$ in terms of the power, $N(f_c, f_1)$, that gets passed through the filter:

$$|H(f)|^2 = \left(\frac{F_s}{2\sigma^2} \right) \frac{dN(f_c, f_1)}{df_1}$$

The power of lowpass noise

Suppose that we have a tone at $f_0 \approx f_c$, and we raise its amplitude, A, until it's just barely audible. The relationship between the tone power and the noise power, at the JND amplitude, is

$$10\log_{10}\left(\frac{N(f_c,f_1)+0.5A^2}{N(f_c,f_1)}\right)=1 \quad\Rightarrow\quad N(f_c,f_1)=\frac{0.5A^2}{10^{1/10}-1}$$

So if we measure the minimum tone amplitude that is audible, as a function of f_c and f_1, then we get

$$|H(f)|^2=\left(\frac{1.93F_s}{\sigma^2}\right)\frac{dA(f_c,f_1)}{df_1}$$

. . . so the shape of the auditory filter is the derivative, with respect to cutoff frequency, of the smallest audible power of the tone at f_c.

Symmetric filters

Using the method on the previous slide, Patterson showed that the auditory filter shape varies somewhat depending on loudness, but the auditory filter centered at f_0 is pretty well approximated as

$$|H(f)|^2 = \frac{1}{(b^2 + (f - f_0)^2)^4}$$

What is the inverse transform of a symmetric filter?

$$|H(f)|^2 = \frac{1}{(b^2 + (f - f_0)^2)^4}$$

Patterson suggested analyzing this as

$$|H(f)|^2 = \frac{1}{(b^2 + (f - f_0)^2)^4} = \left| |G(f)|^2 \right|^4$$

where

$$|G(f)|^2 = \frac{1}{b^2 + (f - f_0)^2}$$

What is the inverse transform of $|G(f)|$?

First, let's just consider the filter

$$|G(f)|^2 = \frac{1}{b^2 + (f - f_0)^2}$$

The only causal filter with this frequency response is the basic second-order resonator filter,

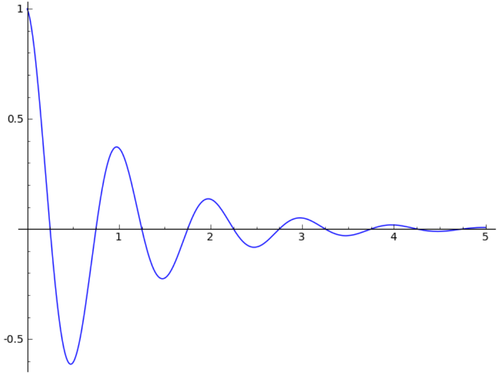
$$G(f) = \frac{2\pi}{2\pi(b - j(f - f_0))}$$

...which is the Fourier transform ($G(f) = \int g(t)e^{-j2\pi ft}dt$) of

$$g(t) = \begin{cases} 2\pi e^{-2\pi(b - jf_0)t} & t > 0 \\ 0 & t < 0 \end{cases}$$

What does $g(t)$ look like?

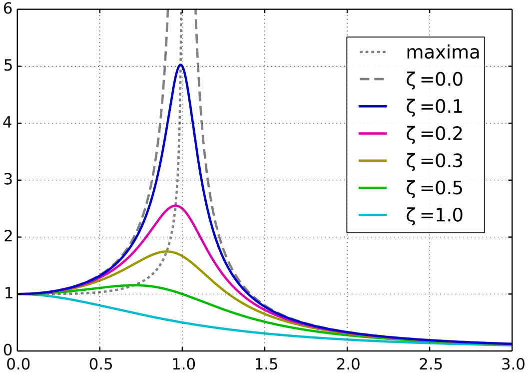
The real part of $g(t)$ is $e^{-2\pi bt}\cos(2\pi f_0 t)u(t)$, which is shown here:



By LM13700, CC-SA3.0, `https://commons.wikimedia.org/wiki/File:DampedSine.png`

What does $|G(f)|$ look like?

$|G(f)|$ looks like this (the frequency response of a standard second-order resonator filter). It's closest to the olive-colored one:



By Geek3, Gnu Free Documentation License,

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

From $G(f)$ to $H(f)$

Patterson suggested analyzing $H(f)$ as

$$|H(f)|^2 = \frac{1}{(b^2+(f-f_0)^2)^4} = \left||G(f)|^2\right|^4$$

which means that

$$h(t) = g(t) * g(t) * g(t) * g(t)$$

So what is $g(t) * g(t)$?

The self-convolution of an exponential is a gamma

Let's first figure out what is $g(t) * g(t)$, where

$$g(t) = e^{-at}u(t), \quad a = 2\pi(b - jf_0)$$

We can write it as

$$\begin{aligned} g(t) * g(t) &= \int_0^t e^{-a\tau} e^{-a(t-\tau)} d\tau \\ &= e^{-at} \int_0^t e^{-a\tau} e^{+a\tau} d\tau \\ &= te^{-at} u(t) \end{aligned}$$

Repeating that process, we get

$$g(t) * g(t) * g(t) * g(t) \propto t^3 e^{-at} u(t)$$

The Gammatone Filter

Patterson proposed that, since $h(t)$ is obviously real-valued, we should model it as

$$H(f) = \left(\frac{1}{b+j(f-f_0)}\right)^n + \left(\frac{1}{b+j(f+f_0)}\right)^n$$

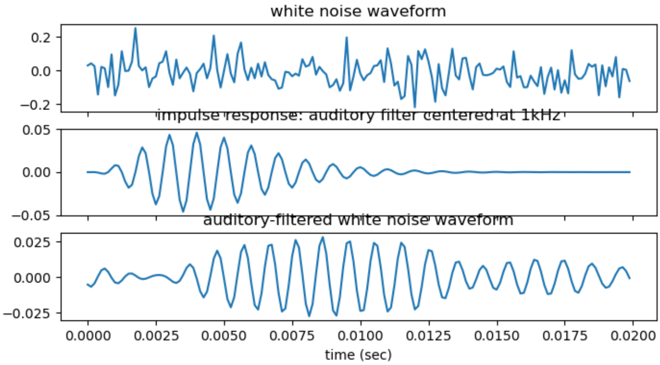
Whose inverse transform is a filter called a **gammatone filter** (because it looks like a gamma function, from statistics, multiplied by a tone):

$$h(t) \propto t^{n-1}e^{-2\pi bt}\cos(2\pi f_0 t)u(t)$$

where, in this case, the order of the gammatone is $n = 4$.

The Gammatone Filter

The top frame is a white noise, $x[n]$. The middle frame is a gammatone filter at $f_c = 1000$Hz, with a bandwidth of $b = 128$Hz. The bottom frame is the filtered noise $y[n]$.



Outline

Summary

- Autocorrelation of filtered noise:

$$r_{yy}[n] = r_{xx}[n] * h[n] * h[-n]$$

- Power spectrum of filtered noise:

$$R_{yy}(\omega) = R_{xx}(\omega)|H(\omega)|^2$$

- Auditory-filtered white noise:

$$E\left[\frac{1}{N}\sum_n y^2[n]\right] = \frac{\sigma^2}{F_s}\int_{-F_s/2}^{F_s/2} |H(f)|^2 df$$

- Bandwidth of the auditory filters:

$$Q = \frac{f}{b} \approx 9.26$$

- Shape of the auditory filters:

$$|H(f)|^2 = \frac{1}{(b^2 + (f - f_0)^2)^4}, \quad h(t) \propto t^{n-1} e^{-2\pi bt} \cos(2\pi f_0 t) u(t)$$