Lecture 3: Noise

Mark Hasegawa-Johnson

ECE 417: Multimedia Signal Processing, Fall 2020

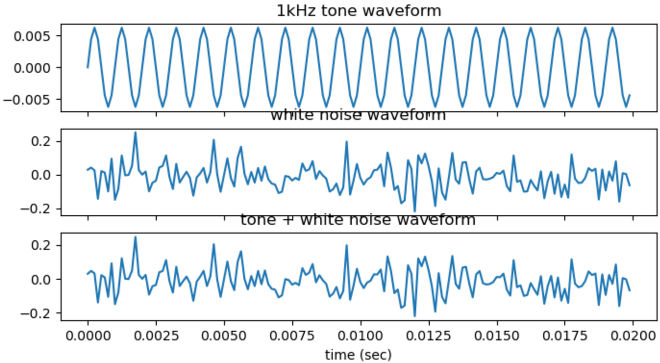
Outline

Noisy Telephones

- In the 1920s, Harvey Fletcher had a problem.
- Telephones were noisy (very noisy).
- Sometimes, people could hear the speech. Sometimes not.
- Fletcher needed to figure out why people could or couldn't hear the speech, and what Western Electric could do about it.

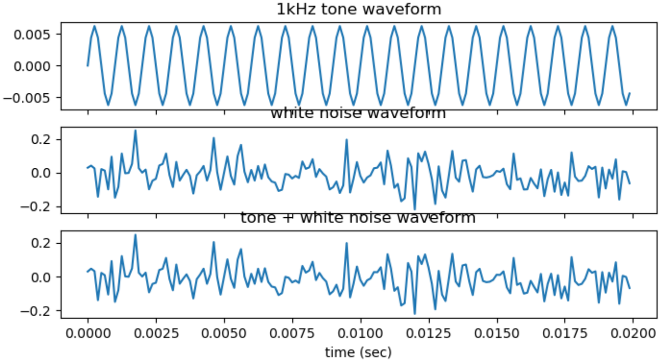
Tone-in-Noise Masking Experiments

He began playing people pure tones mixed with noise, and asking people "do you hear a tone"? If 50% of samples actually contained a tone, and if the listener was right 75% of the time, he considered the tone "audible."



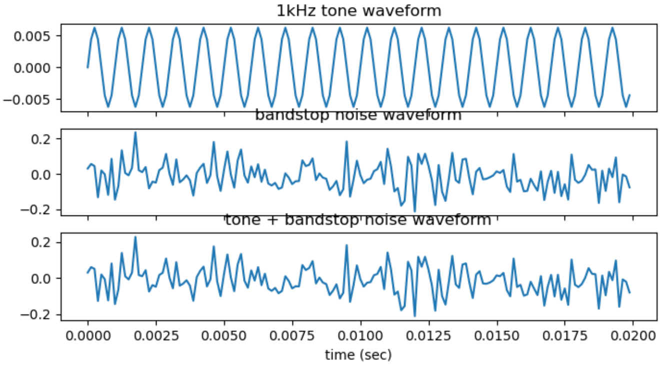
Tone-in-Noise Masking Experiments

People's ears are astoundingly good. This tone is inaudible in this noise. But if the tone was only $2\times$ greater amplitude, it would be audible.

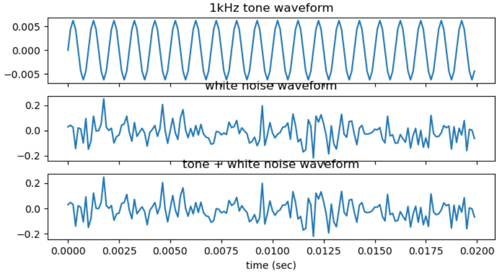


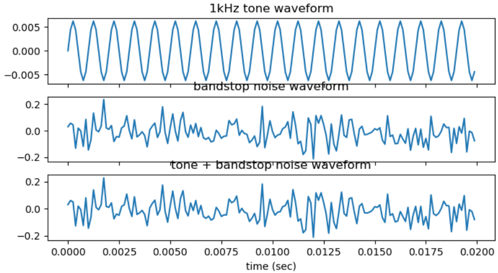
Tone-in-Noise Masking Experiments

Even more astounding: the same tone, in a very slightly different noise, is perfectly audible, to every listener.



What's going on (why can listeners hear the difference?)





Outline

Review: Discrete Fourier Transform

Remember the discrete Fourier transform (DFT):

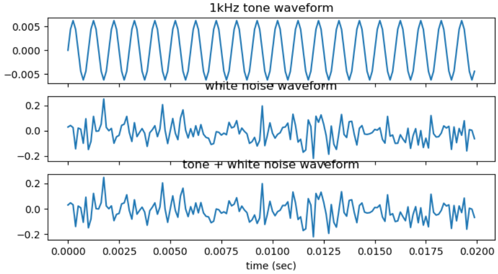
$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi kn}{N}\right)}, \qquad x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k] e^{j\left(\frac{2\pi kn}{N}\right)}$$

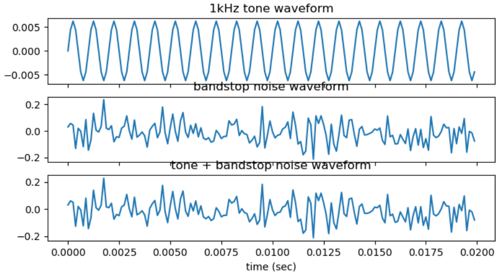
This is useful because, unlike $X(\omega)$, we can actually compute it on a computer (it's discrete in both time and frequency). If $x[n]$ is finite length (nonzero only for $0 \le n \le N-1$), then

$$X[k] = X\left(\omega = \frac{2\pi k}{N}\right)$$

We sometimes write this as $X[k] = X(\omega_k)$, where, obviously, $\omega_k = \frac{2\pi k}{N}$.

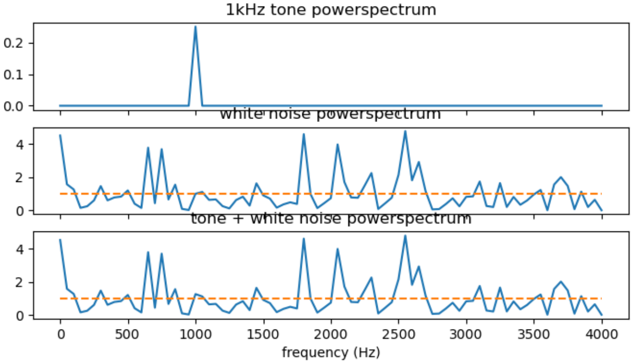
What's going on (why can listeners hear the difference?)





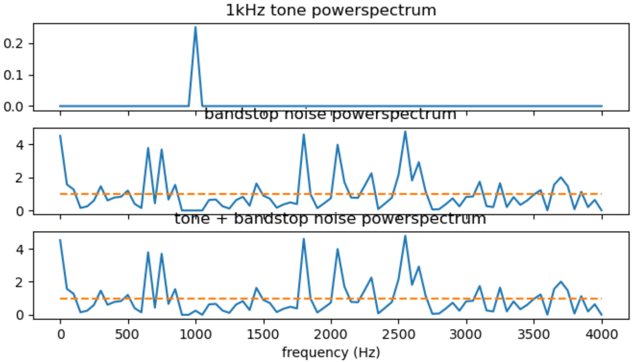
Fourier to the Rescue

Here's the DFT power spectrum ($|X[k]|^2$) of the tone, the white noise, and the combination.



Bandstop Noise

The "bandstop" noise is called "bandstop" because I arbitrarily set its power to zero in a small frequency band centered at 1kHz. Here is the power spectrum. Notice that, when the tone is added to the noise signal, the little bit of extra power makes a noticeable (audible) change, because there is no other power at that particular frequency.



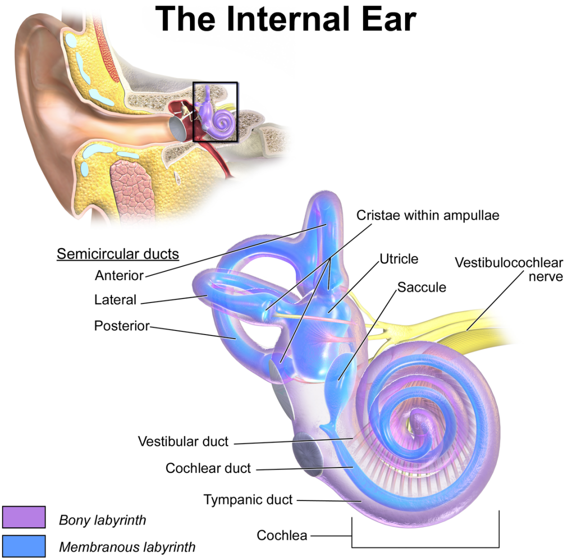
Fletcher's Model of Masking

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 k^{th} bandpass filter, is N_k.
3. The power of the noise+tone is $N_k + T_k$.
4. If there is **any** band, k, in which $\frac{N_k+T_k}{N_k} >$ threshold, then the tone is audible. Otherwise, not.

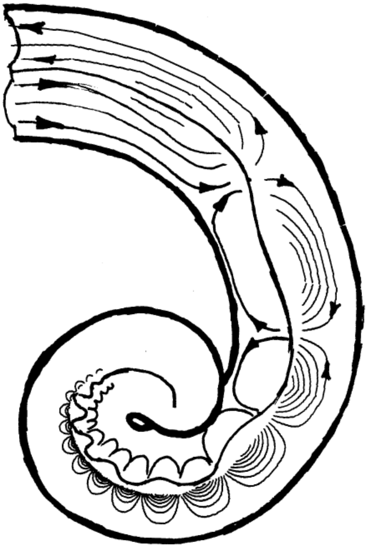
Von Bekesy and the Basilar Membrane

- In 1928, Georg von Békésy found Fletcher's auditory filters.
- Surprise: they are **mechanical**.
- The inner ear contains a long (3cm), thin (1mm), tightly stretched membrane (the basilar membrane). Like a steel drum, it is tuned to different frequencies at different places: the outer end is tuned to high frequencies, the inner end to low frequencies.
- About 30,000 nerve cells lead from the basilar membrane to the brain stem. Each one sends a signal if its part of the basilar membrane vibrates.



Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014." WikiJournal of Medicine 1 (2).

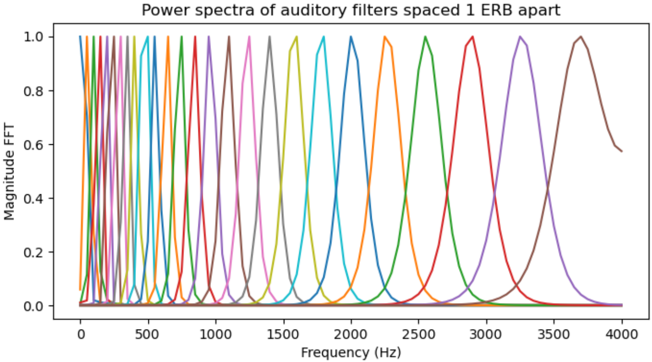
DOI:10.15347/wjm/2014.010. ISSN 2002-4436.



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

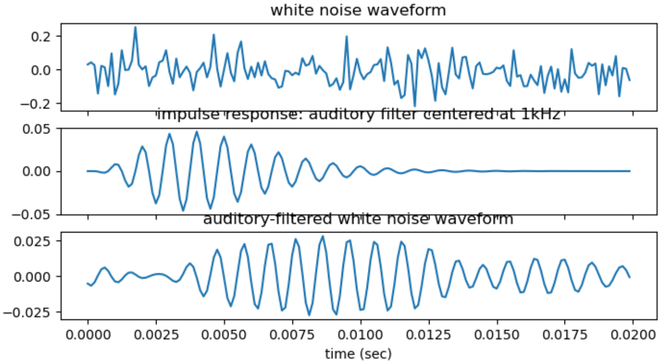
Frequency responses of the auditory filters

Here are the squared magnitude frequency responses ($|H(\omega)|^2$) of 26 of the 30000 auditory filters. I plotted these using the parametric model published by Patterson in 1974:



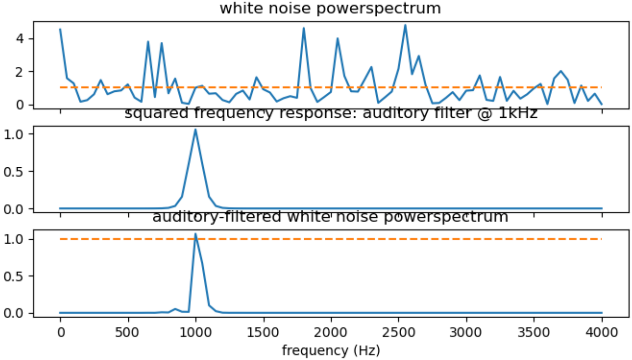
Filtered white noise

An acoustic white noise signal (top), filtered through a spot on the basilar membrane with a particular impulse response (middle), might result in narrowband-noise vibration of the basilar membrane (bottom).



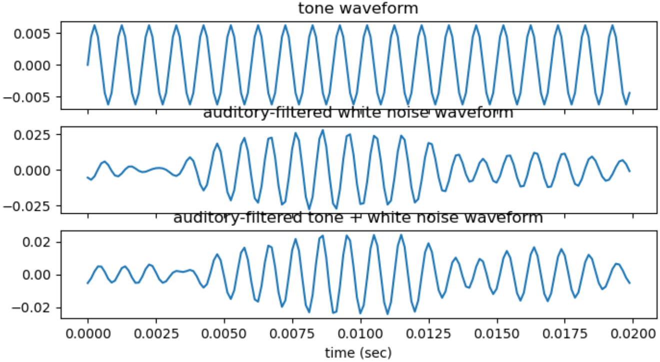
Filtered white noise

An acoustic white noise signal (top), filtered through a spot on the basilar membrane with a particular impulse response (middle), might result in narrowband-noise vibration of the basilar membrane (bottom).



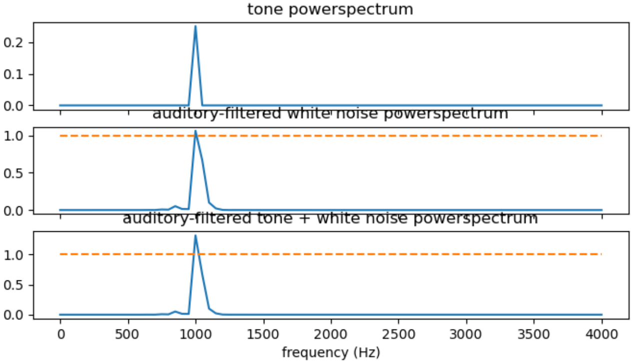
Tone + Noise: Waveform

If there is a tone embedded in the noise, then even after filtering, it's very hard to see that the tone is there. . .



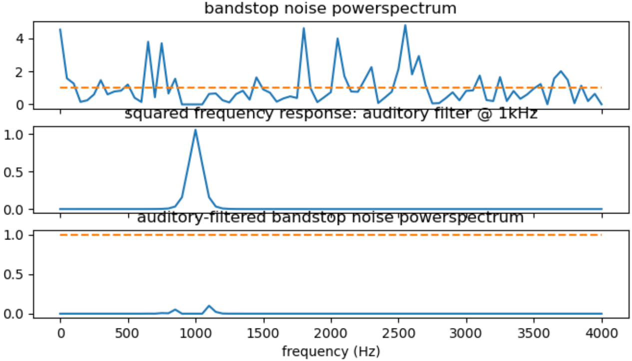
Filtered white noise

But, Fourier comes to the rescue! In the power spectrum, it is almost possible, now, to see that the tone is present in the white noise masker.



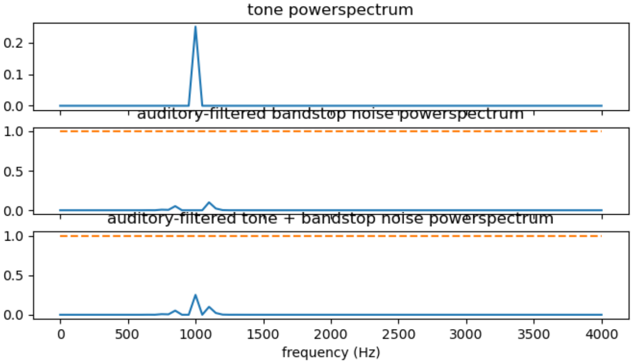
Filtered bandstop noise

If the masker is bandstop noise, instead of white noise, the spectrum after filtering looks very different. . .



Filtered tone + bandstop noise

. . . and the tone+noise looks very, very different from the noise by itself.



This is why the tone is audible!

What an excellent model! Why should I believe it?

Let's spend the rest of today's lecture talking about:

- What is a power spectrum?
- What is noise?
- What is autocorrelation?

Then, next lecture, we will find out what happens to noise when it gets filtered by an auditory filter.

Outline

What is power?

- Power (Watts=Joules/second) is usually the time-domain average of amplitude squared.
- Example: electrical power $P = R\overline{i^2(t)} = \overline{v^2(t)}/R$
- Example: acoustic power $P = \langle z_0 \overline{u^2(t)} \rangle = \overline{p^2(t)}/z_0$
- Example: mechanical power (friction) $P = \mu\overline{v^2(t)} = \overline{f^2(t)}/\mu$

where, by $\overline{x^2(t)}$, I mean the time-domain average of $x^2(t)$.

What is power?

In signal processing, we abstract away from the particular problem, and define instantaneous power as just

$$P = \overline{x^2(t)}$$

or, in discrete time,

$$P = \overline{x^2[n]}$$

Parseval's Theorem for Energy

Parseval's theorem tells us that the energy of a signal is the same in both the time domain and frequency domain. Here's Parseval's theorem for the DTFT:

$$\sum_{n=-\infty}^{\infty} x^2[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2\, d\omega$$

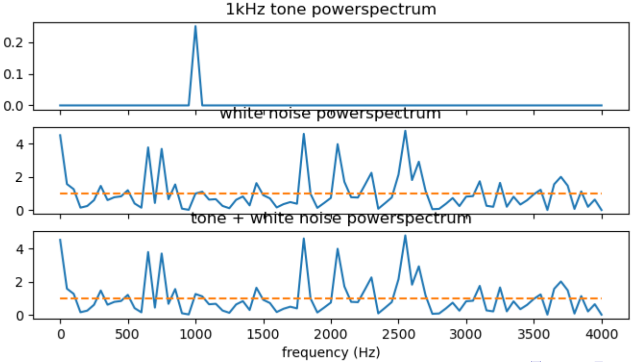
...and here it is for the DFT:

$$\sum_{n=0}^{N-1} x^2[n] = \frac{1}{N}\sum_{k=0}^{N-1} |X[k]|^2$$

Parseval's Theorem

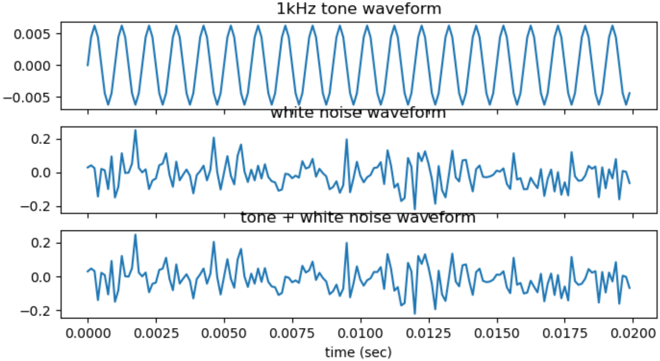
Notice that the white noise spectrum (middle window, here) has an energy of exactly

$$\frac{1}{N}\sum_{k=0}^{N-1}|X[k]|^2 = 1$$



Parseval's Theorem

The window length here is 20ms, at a sampling rate of $F_s = 8000$Hz, so $N = (0.02)(8000) = 160$ samples. The white noise signal is composed of independent Gaussian random variables, with zero mean, and with standard deviation of $\sigma_x = \frac{1}{\sqrt{N}} = 0.079$, so $\sum_{n=0}^{N-1} x^2[n] \approx N\sigma_x^2 = 1$.



Parseval's Theorem for Power

The **Power** of a signal is energy divided by duration. So,

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

. . . and here it is for the DFT:

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = \frac{1}{N^2}\sum_{k=0}^{N-1} |X[k]|^2$$

Power Spectrum

The DFT power spectrum of a signal is defined to be $R[k] = \frac{1}{N}|X[k]|^2$. This is useful because the signal power is

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = \frac{1}{N}\sum_{k=0}^{N-1} R[k]$$

Similary, the DTFT power spectrum of a signal of length N can be defined to be $R(\omega) = \frac{1}{N}|X(\omega)|^2$, because the signal power is

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

In this class we will almost never use the power spectrum of an infinite length signal, but if we need it, it can be defined as

$$R(\omega) = \lim_{N\to\infty} \frac{1}{N}\left|\sum_{n=-(N-1)/2}^{(N-1)/2} x[n]e^{-j\omega n}\right|^2$$

Outline

What is noise?

- "Noise" is a signal, $x[n]$, each of whose samples is a **random variable**.
- For the rest of this course, I'll assume that the noise is **stationary,** which means that the pdf of $x[n]$ is the same as the pdf of $x[n-1]$ (identically distributed).
- If each sample is also **uncorrelated** with the other samples (we write: $x[n] \perp x[n+1]$), we call it **white noise**. This is because (as I will show you soon) its expected power spectrum is flat, like the spectrum of white light.
- The noise we talk about most commonly is **zero-mean Gaussian white noise**, i.e.,

$$x[n] \sim \mathcal{N}(0, \sigma^2), x[n] \perp x[n+1]$$

Sums of Gaussian random variables

Remember that the sum of Gaussian random variables is Gaussian. So any variable z defined as

$$z = a_0 x[0] + a_1 x[1] + \ldots a_{N-1} x[N-1]$$

is itself a Gaussian random variable, with mean given by

$$E[z] = \sum_{n=0}^{N-1} a_n E[x[n]]$$

and with variance given by

$$\sigma_z^2 = \sum_{n=0}^{N-1} a_n^2 \sigma_{x[n]}^2 + \text{(terms that depend on covariances)}$$

In particular, if $x[n]$ is zero-mean Gaussian white noise, then

$$z \sim \mathcal{N}(0, \sum_n a_n^2 \sigma^2)$$

What's the Fourier transform of Noise?

Remember the formula for the DFT:

$$X[k] = \sum_{n=0}^{N-1} e^{-j\omega_k n} x[n], \quad \omega_k = \frac{2\pi k}{N}$$

If $x[n]$ is a zero-mean Gaussian random variable, then so is $X[k]$! More specifically, it is a complex number with Gaussian real and imaginary parts:

$$X_R[k] = \sum_{n=0}^{N-1} \cos(\omega_k n) x[n], \quad X_I[k] = -\sum_{n=0}^{N-1} \sin(\omega_k n) x[n]$$

Using the sums-of-Gaussians formulas on the previous page, you can show that

$$E\left[X_R[k]\right] = E\left[X_R[k]\right] = 0, \quad \mathrm{Var}\left(X_R[k]\right) = \mathrm{Var}\left(X_I[k]\right) = \frac{N\sigma^2}{2}$$

What's the Fourier transform of Noise?

Notice how totally useless it would be to plot the expected value of the DFT — it would always be zero!

$$E\left[X_R[k]\right] = E\left[X_I[k]\right] = 0$$

Instead, it's more useful to plot the variances:

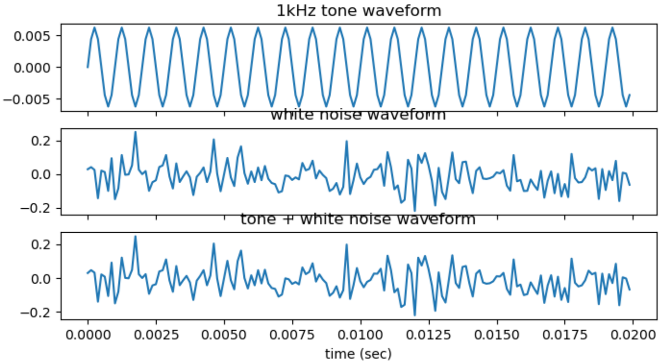
$$\mathrm{Var}\left(X_R[k]\right) = E\left[X_R^2[k]\right] = \frac{N\sigma^2}{2}$$
$$\mathrm{Var}\left(X_I[k]\right) = E\left[X_I^2[k]\right] = \frac{N\sigma^2}{2}$$

In fact, putting those two things together, we get something even nicer:

$$E\left[\frac{1}{N}|X[k]|^2\right] = \frac{1}{N}E\left[X_R^2[k] + X_I^2[k]\right] = \sigma^2$$

An example of White Noise

The window length here is 20ms, at a sampling rate of $F_s = 8000$Hz, so $N = (0.02)(8000) = 160$ samples. The white noise signal is composed of independent Gaussian random variables, with zero mean, and with variance of $\sigma_x^2 = \frac{1}{N}$, so its total energy is $\sum_{n=0}^{N-1} x^2[n] \approx N\sigma^2 = 1$.

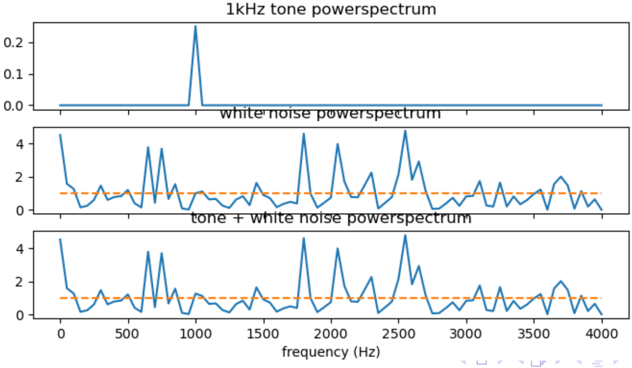


White Noise Energy Spectrum

The energy spectrum $|X[k]|^2$ is itself a random variable, but the expected value of the power spectrum is

$$E\left[|X[k]|^2\right] = E\left[X_R^2[k] + X_I^2[k]\right] = 1$$

which is shown, here, by the dashed horizontal line.



Outline

Inverse DTFT of the Power Spectrum

Since the power spectrum of noise is MUCH more useful than the expected Fourier transform, let's see what the inverse Fourier transform of the power spectrum is. Let's call $R(\omega)$ the power spectrum, and $r[n]$ its inverse DTFT.

$$R(\omega) = \frac{1}{N}|X(\omega)|^2 = \frac{1}{N}X(\omega)X^*(\omega)$$

where $X^*(\omega)$ means complex conjugate. Since multiplying the DTFT means convolution in the time domain, we know that

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

where $z[n]$ is the inverse transform of $X^*(\omega)$ (we haven't figured out what that is, yet).

Inverse DTFT of the Power Spectrum

So what's the inverse DFT of $X^*(\omega)$? If we assume that $x[n]$ is real, we get that

$$
\begin{aligned}
X^*(\omega) &= \left(\sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}\right)^* \\
&= \sum_{n=-\infty}^{\infty} x[n]e^{j\omega n} \\
&= \sum_{m=-\infty}^{\infty} x[-m]e^{-j\omega m}
\end{aligned}
$$

So if $x[n]$ is real, then the inverse DTFT of $X^*(\omega)$ is $x[-n]$!

Autocorrelation

The power spectrum is

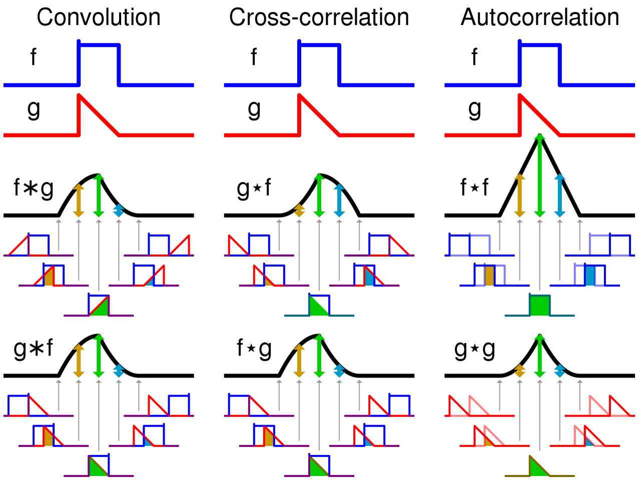
$$R(\omega) = \frac{1}{N}|X(\omega)|^2$$

Its inverse Fourier transform is the autocorrelation,

$$r[n] = \frac{1}{N}x[n] * x[-n] = \frac{1}{N}\sum_{m=-\infty}^{\infty} x[m]x[m-n]$$

This relationship, $r[n] \leftrightarrow R(\omega)$, is called Wiener's theorem, named after Norbert Wiener, the inventor of cybernetics.

Convolution vs. Autocorrelation



Autocorrelation is also a random variable!

- Notice that, just as the power spectrum is a random variable, the autocorrelation is also a random variable.
- The autocorrelation is the average of N consecutive products, thus

$$E\left[r[n]\right] = E\left[\frac{1}{N}\sum_{m=0}^{N-1} x[m]x[m-n]\right] = E\left[x[m]x[m-n]\right]$$

  ...where the last form only makes sense if the signal is stationary (all samples identically distributed), so that $E\left[x[m]x[m-n]\right]$ doesn't depend on m.
- The expected autocorrelation is related to the covariance and the mean:

$$E\left[r[n]\right] = \text{Cov}\left(x[m], x[m-n]\right) + E\left[x[m]\right] E\left[x[m-n]\right]$$

- If $x[n]$ is zero-mean, that means

$$E\left[r[n]\right] = \text{Cov}\left(x[m], x[m-n]\right)$$

Autocorrelation of white noise

If $x[n]$ is zero-mean white noise, then

$$E\left[r[n]\right] = E\left[x[m]x[m-n]\right] = \begin{cases} \sigma^2 & n = 0 \\ 0 & \text{otherwise} \end{cases}$$

We can write

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

Outline

Summary

- 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.
- **Parseval's Theorem:**

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

- **Wiener's Theorem:**

$$R(\omega) \leftrightarrow r[n] = \frac{1}{N}x[n] * x[-n]$$

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

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